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

Anonymous authors
Paper under double-blind review

Abstract

The subdominant (minmax) ultrametric is a canonical tree-structured summary of a dissimilarity matrix, arising equivalently as the ultrametric induced by single-linkage clustering. While its classical stability theory is usually formulated in ℓ_∞ or Gromov–Hausdorff terms, such bounds are poorly suited to sparse perturbations that alter only a few pairwise distances. We develop an ℓ_0-type stability theory for this operator. Our analysis shows that sparse edits propagate only through the *minimum spanning tree* (MST): a pairwise ultrametric value can change only if its tree path crosses an edited edge or a cut newly exposed by an edited off-tree edge. This yields a sharp per-edit exposed-cut score and a tree-only global envelope, leading to Hamming–Lipschitz bounds on the number of ultrametric entries that can change. We also prove sharpness results showing that this dependence on tree geometry is unavoidable: under strict cut separation the tree-edge bound is attained exactly, and for off-tree edits there are explicit families in which one edited distance changes $\Theta(n^2)$ ultrametric entries. In addition, we prove a conditional near-additivity principle for multiple edits under certified large per-edit changed regions and negligible aggregate overlap. Experiments on deep-embedding graphs show that the resulting structural scores provide useful vulnerability diagnostics for hierarchical representations.

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 is maximally dominated by the given dissimilarity and plays a canonical role in hierarchical clustering and metric geometry. 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 maps a dissimilarity to its associated subdominant ultrametric, equivalently the merge-height function of single-linkage clustering. 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 ultrametric 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 ℓ_0-type theory in which the changed-pair set is controlled by per-edit exposed-cut regions. In particular, the analysis yields a tree-only global envelope $\bar{L}_T$, giving bounds of the form $\|u_d - u_{\tilde{d}}\|_0 \leq \bar{L}_T \, \|d - \tilde{d}\|_0$. $\bar{L}_T$ depends only on the MST structure; in the worst case on an n-node graph, $\bar{L}_T \leqslant \binom{n}{2}$. 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.	$\|u_d - u_{\tilde{d}}\|_0 \leq \bar{L}_T \, \|d - \tilde{d}\|_0$	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.** For each edited pair, we define the sharp per-edit affected size, which localizes the pairs that can change through the cuts actually exposed by that edit (Theorem 2). We also introduce a tree-only envelope $\bar{L}_T$, obtained by maximizing the union of cut-rectangles along MST paths. This yields a structural bound on how many entries of u_d can move under a sparse perturbation, while cleanly separating the edit-dependent sharp quantity from the tree-only global one.

- **Sharpness and worst-case behavior.** We show that our bound is attained exactly for tree-edge edits under strict cut separation and is attained on explicit off-tree families (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 multi-edit perturbations, we give a conditional near-additivity principle: if one can certify per-edit changed regions that nearly saturate the exposed-region scores and have negligible aggregate overlap, then the total number of changed ultrametric entries is asymptotically close to the sum of those scores (Corollary 1).
- **Simple case studies as diagnostics.** Finally, we use the theorem-motivated structural score $S_{\text{union}}(e) = |A_e||B_e|$ 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 the theory can inform practical vulnerability maps even though the experiments are diagnostic rather than task-optimized.

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

2 Notation, 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). Because $\|\cdot\|_0$ counts upper-triangular matrix entries, all Hamming counts in this paper are over unordered pairs. 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 $e = \{i,j\} \in E$. We define the set of all unordered pairs as $\binom{V}{2} := \big\{\{i,j\} : i,j \in V,\ i \neq j\big\}$. Let $T = (V, E(T))$ denote the unique MST of d on G (we assume a strict lexicographic tie-breaking rule to guarantee uniqueness; see Assumption 1 for details). 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 can then define the associated cut-pair set $\mathcal{R}_e := \big\{\{i,j\} \in \binom{V}{2} : i \in A_e,\ j \in B_e\big\}$. Thus $\mathcal{R}_e$ is the unordered-pair analogue of the rectangle $A_e \times B_e$.

We abbreviate $w_e := d(e)$ and define the alternative cut minimum

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

and the corresponding cut gap

$$\Delta_e(d) := w_e^+(d) - w_e \geq 0. \tag{3}$$

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 $(w(e), \operatorname{rank}_\prec(e))$. Equivalently, conceptually perturb*

$$w'(e) \;=\; w(e) \;+\; \eta \operatorname{rank}_\prec(e), \qquad \textit{with a single infinitesimal } \eta > 0, \tag{4}$$

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. This infinitesimal perturbation is used only to select a unique MST and a deterministic processing order among equal-weight edges; all quantities $d(e)$, u_d, $w_e^+(d)$, and $\Delta_e(d)$ refer to the original unperturbed dissimilarity values unless explicitly stated otherwise.

Under Assumption 1, the MST is unique and its tree edges are strictly ordered under the lexicographically perturbed weights. We next recall the minmax ultrametric and the standard MST facts used in our analysis. In general, $\Delta_e(d)$ need not be strictly positive: lexicographic tie-breaking guarantees a unique MST and a deterministic edge order, but it does not alter the original numeric edge weights. We say that a tree edge $e \in E(T)$ is strictly cut-separated if $\Delta_e(d) > 0$.

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 ℓ_0 sense:

$$\|d - \tilde{d}\|_0 \;:=\; \#\big\{\, \{i,j\} \subseteq S : i < j,\; d(i,j) \neq \tilde{d}(i,j) \,\big\}. \tag{5}$$

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

$$U : (X, d) \mapsto (X, u_d). \tag{6}$$

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

$$U(d) = u_d, \qquad u_d(i,j) \text{ for the } (i,j)\text{-entry of the ultrametric.} \tag{7}$$

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.

We now formally define the minmax ultrametric below:

Definition 1 (Minmax subdominant 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. Explicitly, for any $i, j \in S$ with $i \neq j$, it is constructed by minimizing the bottleneck over all possible paths between i and j:*

$$u_d(i,j) = \min_{P \in \mathcal{P}(i,j)} \max_{1 \leqslant t \leqslant m} d(v_{t-1}, v_t) \tag{8}$$

where $\mathcal{P}(i,j)$ denotes the set of all finite paths $P = (v_0, v_1, \ldots, v_m)$ from $v_0 = i$ to $v_m = j$.

2.1 Two MST Facts Used Throughout

Our analysis repeatedly uses two standard facts about the minimum spanning tree T of d. These are common Lemmas on MST, one can easily find them in the standard algorithms textbook by Cormen et al. (2022).

Lemma 1 (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 2 (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 principle for the subdominant ultrametric under sparse edge edits. At the level of the MST, it shows that if the tree path between two points contains no edited tree edges, then their ultrametric value cannot increase, and it remains unchanged whenever no edited off-tree edge creates a cheaper crossing across any fundamental cut along that path. Globally, the theorem identifies a set of potentially affected tree edges and proves that every changed ultrametric entry must lie in the union of the associated cut-pair sets. This yields an explicit combinatorial upper bound on the support of $u_d - u_{\tilde{d}}$ in terms of the forest obtained by deleting those affected edges from the MST. For single-linkage clustering, the interpretation is that sparse perturbations can only propagate through edited or newly exposed cuts: subtrees separated from these cuts remain rigid, while only the corresponding branches of the hierarchy can move. Thus, Theorem 1 gives 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 $d : \binom{V}{2} \mapsto \mathbb{R}_{\geqslant 0}$, and let $T = (V, E(T))$ be the (tie-broken under Assumption 1) MST of d. 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. By Lemma 1, $u_d(i,j) = \max_{e \in P_T(i,j)} w_e$, and therefore 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 $u_{\tilde{d}}(i,j) = u_d(i,j)$.

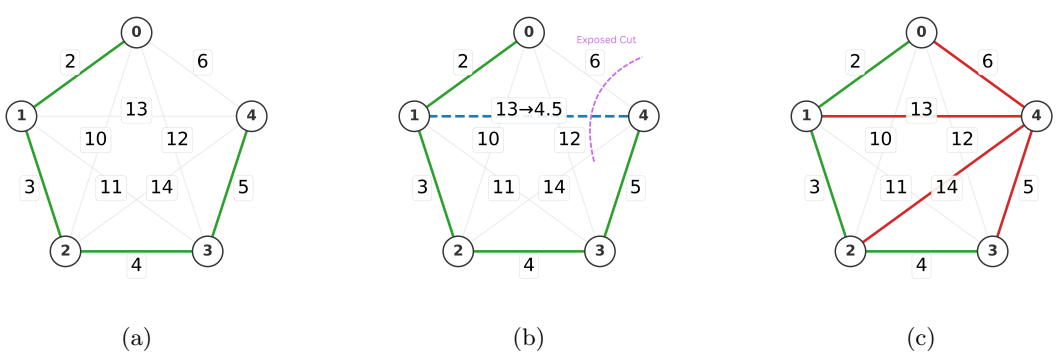
(iii) (Pair-count bound for possible changes) Define the set of potentially affected MST edges

$$\mathcal{E} := \big\{ e \in E(T) : e \in F \text{ or } \exists f \in F \text{ crossing } C_e \text{ with } \tilde{d}(f) < w_e \big\}. \tag{9}$$

Then the number of unordered pairs whose ultrametric value changes satisfies

$$\Big| \big\{ \{i,j\} \in \tbinom{V}{2} : u_{\tilde{d}}(i,j) \neq u_d(i,j) \big\} \Big| = \|u_d - u_{\tilde{d}}\|_0 \leq \Big| \bigcup_{e \in \mathcal{E}} \mathcal{R}_e \Big| = \binom{n}{2} - \sum_{t=1}^{m} \binom{|C_t|}{2}, \tag{10}$$

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-pair set $\mathcal{R}_{\{4,5\}} = \big\{\{i,4\} : i \in \{0,1,2,3\}\big\}$, upper-bounds the total number of changed ultrametric entries, i.e. $\|u_d - u_{\tilde{d}}\|_0$.

Remark. *Theorem 1 reveals an important asymmetry. If the tree path $P_T(i,j)$ contains no edited tree edge, then the original tree bottleneck remains available under $\tilde{d}$, so the ultrametric value $u_{\tilde{d}}(i,j)$ can never exceed $u_d(i,j)$. In that regime, a change can only occur through a strictly cheaper off-tree crossing of one of the fundamental cuts along $P_T(i,j)$. Thus sparse perturbations act in two qualitatively different ways: edits on the tree path can raise or lower merge heights directly, whereas off-tree edits can only lower them by opening cheaper alternative connections across MST cuts.*

Also, it gives a first coarse localization result: it shows that all changes in the ultrametric are confined to a union of cut-pair sets associated with a small set of MST edges determined by the edit pattern. However, this description is still global across edits and does not yet yield a Hamming–Lipschitz inequality that separates the contribution of individual edited pairs. To sharpen the picture, we pass to the level of a single edited pair $f = \{x, y\}$. For each such f, we isolate the subset of tree edges whose cuts are actually exposed by that edit; these are the exposed cuts $\Xi(f)$. Formally, for an edited pair $f = \{x, y\} \in F$, define the set of exposed cuts by $\Xi(f) := (\{f\} \cap E(T)) \cup \{e \in E(T) : f \text{ crosses } C_e \text{ and } \tilde{d}(f) < w_e\}$. The next theorem defines the sharp per-edit exposed-cut score

$$S_{\text{union}}(f) := \left| \bigcup_{e \in \Xi(f)} \mathcal{R}_e \right|, \tag{11}$$

and shows that every ultrametric entry changed by the edit must lie inside this exposed region. By further enlarging the exposed cuts to the full tree path $P_T(x, y)$, we obtain a tree-only global envelope $\bar{L}_T$, which yields a Hamming–Lipschitz type bound for arbitrary sparse perturbations.

Theorem 2 (Hamming–Lipschitz bound via exposed cuts)**.** *Let $d : \binom{V}{2} \to \mathbb{R}_{\geq 0}$ be a dissimilarity on a finite set V, and let $T = (V, E(T))$ be the (tie-broken under Assumption 1) MST of d. Let $\tilde{d}$ be any perturbed dissimilarity, and let $F := \{f \in \binom{V}{2} : \tilde{d}(f) \neq d(f)\}$ be its edit support. For an edited pair $f = \{x, y\} \in F$, define the set of exposed cuts by*

$$\Xi(f) := (\{f\} \cap E(T)) \cup \{e \in E(T) : f \text{ crosses } C_e \text{ and } \tilde{d}(f) < w_e\}. \tag{12}$$

Let $C := \{\{i, j\} \in \binom{V}{2} : u_{\tilde{d}}(i, j) \neq u_d(i, j)\}$ denote the set of unordered pairs whose ultrametric values change. Further, define the sharp per-edit exposed-cut size $S_{\text{union}}(f) := \left|\bigcup_{e \in \Xi(f)} \mathcal{R}_e\right|$. For an edited pair $f = \{x, y\}$, define the tree-only path envelope $\bar{S}_T(f) := \left|\bigcup_{e \in P_T(x,y)} \mathcal{R}_e\right|$, and the global tree-only constant $\bar{L}_T := \max_{f \in \binom{V}{2}} \bar{S}_T(f)$.
Then:

(i) $C \subseteq \bigcup_{f \in F} \bigcup_{e \in \Xi(f)} \mathcal{R}_e$. Consequently, $\|u_d - u_{\tilde{d}}\|_0 = |C| \leq \left|\bigcup_{f \in F} \bigcup_{e \in \Xi(f)} \mathcal{R}_e\right|$.

(ii) for every edited pair $f \in F$,

$$S_{\text{union}}(f) \leq \bar{S}_T(f) \leq \bar{L}_T, \tag{13}$$

and therefore

$$\|u_d - u_{\tilde{d}}\|_0 \leq \sum_{f \in F} S_{\text{union}}(f) \leq \sum_{f \in F} \bar{S}_T(f) \leq \bar{L}_T \, |F| = \bar{L}_T \, \|d - \tilde{d}\|_0. \tag{14}$$

In particular,

$$\bar{L}_T \leq \binom{|V|}{2}. \tag{15}$$

Remark. *Theorem 2 separates two distinct layers of control. The quantity $S_{\text{union}}(f)$ is the sharp, edit-dependent score: it records only those MST cuts that are actually exposed by the specific perturbation of f. In contrast, $\bar{S}_T(f)$ and $\bar{L}_T$ are tree-only envelopes obtained by enlarging the exposed region to the full MST path of the edited pair. Thus the theorem distinguishes the mechanism of propagation, which is localized through exposed cuts, from a coarser but globally uniform capacity of the tree to transmit sparse perturbations.*

It yields a Hamming–Lipschitz upper bound that is both localized and structurally interpretable: sparse perturbations can propagate only through exposed MST cuts, and their total effect is controlled by the associated union of cut-pair sets. This naturally raises the next question: are these quantities merely proof-level upper bounds, or do they capture the true scale of instability of the subdominant ultrametric? In particular, one would like to know whether the sharp per-edit score $S_{\mathrm{union}}(f)$ can actually be attained, and whether the resulting dependence on tree geometry is intrinsic. Theorem 3 answers this by showing exact attainability for tree-edge edits under strict cut separation, and sharpness on explicit off-tree families, including worst-case examples in which a single sparse edit changes $\Theta(n^2)$ ultrametric entries.

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 an edited pair $f = \{x, y\}$ and edited dissimilarity $\tilde{d}$ with $\tilde{d}(g) = d(g)$ for $g \notin \{f\}$, define*

$$\Xi(f) \;:=\; (\{f\} \cap E(T)) \cup \{\, e \in E(T) : \; f \text{ crosses } C_e \text{ and } \tilde{d}(f) < w_e \,\}, \qquad S_{\mathrm{union}}(f) := \left| \bigcup_{e \in \Xi(f)} \mathcal{R}_e \right|. \tag{16}$$

Then:

(i) (Tree-edge edit under strict cut separation) *If $f = e \in E(T)$ and e is strictly cut-separated, i.e. $\Delta_e(d) = w_e^+(d) - w_e > 0$, then there exists $\tilde{d}$ supported on $\{f\}$ such that*

$$\|u_d - u_{\tilde{d}}\|_0 = |A_e||B_e| = S_{\mathrm{union}}(f). \tag{17}$$

(ii) (Off-tree sharpness on an explicit family) *There exist dissimilarities d, off-tree edges $f \notin E(T)$, and single-edge edits $\tilde{d}$ supported on $\{f\}$ such that*

$$\|u_d - u_{\tilde{d}}\|_0 = \left| \bigcup_{e \in \Xi(f)} \mathcal{R}_e \right| = S_{\mathrm{union}}(f) = \Theta(n^2). \tag{18}$$

Thus, the upper bound of Theorem 2 is attained on explicit off-tree instances, and in the worst case a single off-tree edit can force a quadratic number of ultrametric changes.

(iii) (Necessity of instance dependence) *Consequently, no universal subquadratic function $c(n) = o(n^2)$ can satisfy*

$$\|u_d - u_{\tilde{d}}\|_0 \leq c(n)\, \|d - \tilde{d}\|_0 \tag{19}$$

for all instances.

Remark. *Theorem 3 shows that the exposed-cut quantities from Theorem 2 are not artifacts of the proof. In particular, the instability of the subdominant ultrametric is governed by the geometry of the MST itself: a single sparse edit can trigger changes on the scale of an entire cut-pair set, and in explicit off-tree families this propagation reaches quadratic size. Thus the dependence on tree geometry is intrinsic to the operator, rather than a byproduct of our bounding technique.*

Theorem 3 resolves the single-edit case: it shows that the exposed-cut score can be attained exactly in structured settings and that, in the worst case, even one edited distance can induce $\Theta(n^2)$ ultrametric changes. *The next natural question is how multiple sparse edits interact.* If their exposed regions substantially overlap, the total number of changed entries may be far smaller than the sum of the individual scores; if they are largely disjoint, one expects an approximately additive effect. Corollary 1 formalizes this latter regime in a conditional form: whenever one can certify large per-edit changed regions with negligible aggregate overlap, the overall Hamming change is asymptotically close to the sum of the corresponding exposed-cut scores.

Corollary 1. *Fix a minimum spanning tree T of d, an edit set $F \subseteq \binom{V}{2}$, and edited weights $\tilde{d}$ supported on F. For each $f \in F$, define the exposed region $R(f) := \bigcup_{e \in \Xi(f)} \mathcal{R}_e$, so that $|R(f)| = S_{\text{union}}(f)$. Assume that for each $f \in F$ there exists a certified changed-pair set $Q(f) \subseteq R(f)$ such that every pair in $Q(f)$ indeed changes under the common edited dissimilarity $\tilde{d}$, i.e. $Q(f) \subseteq \{\{i,j\} \in \binom{V}{2} : u_{\tilde{d}}(i,j) \neq u_d(i,j)\}$. Then:*

(i)

$$\left| \bigcup_{f \in F} Q(f) \right| \leq \|u_d - u_{\tilde{d}}\|_0 \leq \left| \bigcup_{f \in F} R(f) \right| \leq \sum_{f \in F} S_{\text{union}}(f).$$

(ii) Moreover, consider any asymptotic regime of instances (for example, $|V| = n \to \infty$) in which

$$\sum_{f \in F} |Q(f)| = (1 - o(1)) \sum_{f \in F} S_{\text{union}}(f),$$

and the certified regions have asymptotically negligible total overlap:

$$\sum_{\substack{f, f' \in F \\ f < f'}} |Q(f) \cap Q(f')| = o\left(\sum_{f \in F} |Q(f)| \right).$$

Then

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

Remark. *Corollary 1 isolates two logically distinct tasks. The first is a* geometric *task: identify exposed regions $R(f)$ that contain all pairs that could possibly change. The second is a* certification *task: exhibit subsets $Q(f) \subseteq R(f)$ whose pairs are guaranteed to change under the common perturbation. Once such certified regions are available, the global Hamming count is governed purely by set overlap. In this sense, the corollary turns the multi-edit problem from one of ultrametric analysis into one of certifiable combinatorial packing.*

Corollary 1 completes the picture by showing how the single-edit geometry from Theorems 1–3 extends to sparse multi-edit regimes. The key issue is no longer the effect of an individual edit, but the interaction among their exposed regions. When these regions overlap heavily, different edits compete for the same ultrametric entries and the total effect can be far smaller than the sum of their individual scores. When one can instead certify large per-edit changed regions with negligible aggregate overlap, the total Hamming change becomes asymptotically additive. Thus the corollary identifies the precise structural mechanism behind near-additivity: it is not sparsity alone, but sparsity together with low-overlap exposure in the MST geometry.

4 Empirical Case Study: Vulnerability Maps of Deep Embeddings

Our contribution is primarily theoretical, so the role of this section is diagnostic rather than benchmark-driven. Theorems 1 and 2 show that sparse metric edits can change the ultrametric only through edited or newly exposed MST cuts, and that the extent of propagation is controlled by the associated cut-pair sets. This suggests a concrete empirical question: *in representation graphs arising from real data, are there only a few load-bearing MST edges whose perturbation causes substantial ultrametric damage, while most edges are comparatively harmless?* The experiment below is designed to probe exactly that question.

Motivation from the theory. For a tree edge $e \in E(T)$, Theorem 2 reduces the per-edit exposed region to the single cut-rectangle associated with e, whose cardinality is $S_{\text{union}}(e) = |A_e|\,|B_e|$. Thus $S_{\text{union}}(e)$ is the natural structural score predicted by the theory: it quantifies how many unordered pairs can potentially be

affected when the hierarchy is stressed across that cut. Our main empirical goal is therefore to test whether this score meaningfully ranks edges by realized vulnerability in real representation graphs.

4.1 Protocol

We use three 10-class image datasets: CIFAR-10, ImageNet-10, and STL-10. For each dataset, we start from precomputed DINO-ViT features, apply UMAP to obtain a 20-dimensional embedding, and subsample a few thousand points for efficiency. On each embedding, we build the complete Euclidean graph, compute its tie-broken MST T, and obtain the induced subdominant ultrametric u_d via the MST bottleneck formula. We then compute the structural score $S_{\text{union}}(e) = |A_e||B_e|$ for every tree edge $e \in E(T)$. Theorem 1 predicts that ultrametric changes must remain localized to unions of such cut-rectangles, while Theorem 2 identifies $S_{\text{union}}(e)$ as the sharp tree-edge quantity controlling the size of the exposed region. Figure 2 therefore serves two purposes: the bottom row visualizes the empirical distribution of this theorem-motivated score across the MST, and the top row evaluates whether targeting high-score edges indeed produces larger realized ultrametric damage.

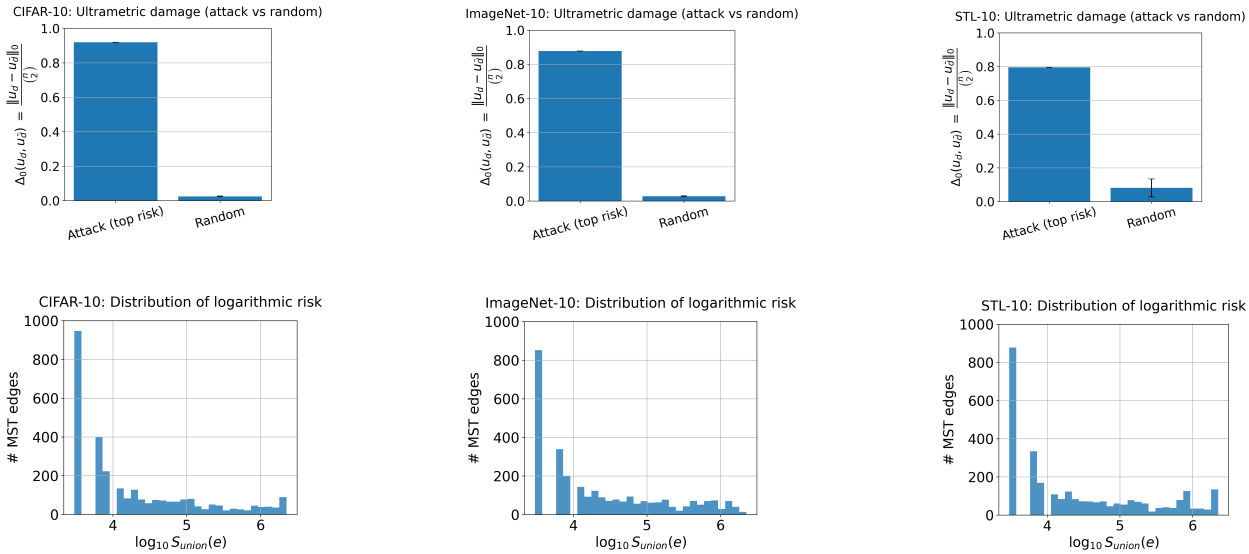


Figure 2: **Top row:** normalized ultrametric damage $\Delta_0(u_d, u_{\tilde{d}})$ under targeted (top-risk) versus random sparse perturbations for CIFAR-10, ImageNet-10, and STL-10. **Bottom row:** histograms of the theorem-motivated structural score $\log_{10} S_{\text{union}}(e)$ over MST edges for the same three datasets. It shows that structural risk is concentrated in a small high-score tail, while the top row shows that targeting that tail causes substantially larger realized Hamming damage than random edits with the same budget, consistent with Theorems 1 and 2.

To test this, we fix a sparse edit budget m equal to 10% of the MST edges and compare two strategies:

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

In both cases, we construct a perturbed metric $\tilde{d}$ by sharply inflating the selected tree-edge weights, recompute the MST and the induced ultrametric $u_{\tilde{d}}$, and measure the normalized Hamming distortion

$$\Delta_0(u_d, u_{\tilde{d}}) := \frac{\|u_d - u_{\tilde{d}}\|_0}{\binom{n}{2}}. \tag{20}$$

This is a stress test inspired by the tree-edge exposed-cut analysis. It is not intended as a literal validation of the sharpness theorem edge-by-edge, but rather as an empirical probe of whether the score from Theorem 2 tracks practical vulnerability.

4.2 Results

Figure 2 (bottom row) shows that the distribution of $\log_{10} S_{\text{union}}(e)$ is strongly skewed for all three datasets. Most MST edges lie in a low-score bulk, corresponding to leaf-like or weakly load-bearing cuts, while a small number of bridge-like edges form a thin high-score tail. This is precisely the type of heterogeneity suggested by Theorems 1 and 2: sparse perturbations should not propagate uniformly through the hierarchy, but instead concentrate around a small set of structurally important cuts.

Figure 2 (top row) shows that this structural heterogeneity is not merely combinatorial. Across CIFAR-10, ImageNet-10, and STL-10, the top-risk attack consistently induces much larger normalized Hamming damage than the random baseline at the same edit budget. In particular, even in budgets where random edits leave $\Delta_0(u_d, u_{\tilde{d}})$ close to zero, editing the highest-score edges already flips a visible fraction of ultrametric entries. The practical conclusion is that the theorem-motivated score $S_{\text{union}}(e)$ identifies a small set of load-bearing edges whose perturbation produces disproportionate global impact.

Overall, this case study should be read as a diagnostic interpretation of the theory. Theorem 1 explains *where* propagation can occur, while Theorem 2 supplies the natural per-edge structural score. Figure 2 shows that, on real deep-embedding graphs, these quantities yield a meaningful vulnerability map: most MST edges are structurally benign, but a small high-score tail captures the edges whose sparse perturbation can ripple widely through the induced hierarchy. Taken together, these experiments provide an empirical counterpart to the theoretical picture: the structural score $S_{\text{union}}(e)$ identifies a small set of load-bearing MST edges whose perturbation has disproportionate global impact, while most edges are comparatively benign.

NOTE: *Beyond the deep embedding vulnerability maps presented above, we also evaluate the practical utility of the structural score in a downstream task. In Appendix C, we provide an additional empirical case study on MST-based superpixel image segmentation. This supplementary experiment demonstrates that restricting sparse edits to edges with low $S_{union}(e)$ scores preserves the downstream clustering structure significantly better than baseline heuristics relying solely on local edge weights.*

5 Conclusion

We developed a sparsity-aware stability theory for the subdominant (minmax) ultrametric, complementing classical ℓ_∞/Gromov–Hausdorff results with an ℓ_0-type 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 localization in terms of exposed cut-pair sets, a sharp per-edit score $S_{\text{union}}(f)$, and a tree-only global envelope $\bar{L}_T$. We further proved that this instance dependence is unavoidable by exhibiting single-edit constructions with $\Theta(n^2)$ changes, and identified explicit single-edit sharpness examples together with conditional multi-edit regimes in which the upper bound is asymptotically 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

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz, et al. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, pp. 3–4, 2013.

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.

Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. 2022.

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.

Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2): 179–188, 1936.

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.

C Kaynak. Methods of combining multiple classifiers and their applications to handwritten digit recognition. *Unpublished master's thesis, Bogazici University*, 1995.

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.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 2002.

Ling Liu, Yanzhao Wu, Wenqi Wei, Wenqi Cao, Semih Sahin, and Qi Zhang. Benchmarking Deep Learning Frameworks: Design Considerations, Metrics and Beyond. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, pp. 1258–1269, July 2018. doi: 10.1109/ICDCS.2018.00125.

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

James B McQueen. Some methods of classification and analysis of multivariate observations. In *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.*, pp. 281–297, 1967.

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.

Charles T Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on computers*, 100(1):68–86, 2006.

Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. *Advances in neural information processing systems*, 17, 2004.

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.