LATENT DISTANCE DENOISING IN METRIC MEASURE SPACES

Anonymous authors
Paper under double-blind review

ABSTRACT

Pairwise observations often encode a noisy monotone transform of an unobserved distance. We study how to denoise all latent pairwise distances when the underlying points are sampled from a metric measure space governed only by lower volume-growth bounds, without assuming coordinates, curvature, or a fixed dimension. Our observable neighborhood construction converts internal similarity into controlled-radius clusters and then denoises distances by independent cluster-to-cluster averaging. At any fixed target accuracy, the resulting algorithm produces uniform latent-distance estimates in near-linear time up to polylogarithmic factors. Controlled simulations reduce average error by 86% on an interval and 77% on a branching metric tree. Matching lower-bound and hardness results clarify the remaining statistical and computational limitations.

1 INTRODUCTION

Relational data reveal interactions while hiding the distances that generated them. We assign each object a latent point X_v in a metric space with distance d and let its pairwise observation depend only on $d(X_u, X_v)$. Our target is the finite latent-distance matrix, not coordinates or a Euclidean realization. Unlike smooth geometric inference, which can exploit charts, curvature, and regular geodesics, we allow trees, stratified spaces, fractals, rough subsets, and finite metrics. Metric measure spaces provide a common intrinsic language for these examples (Heinonen, 2001; Sturm, 2006; Villani, 2009), but no local coordinates.

This distinction changes the reconstruction problem. A coordinate-based method may first estimate positions and then read off distances, whereas an intrinsic method must compare hidden neighborhoods using only relations among sampled vertices. Consequently, even defining a reliable local average becomes part of the statistical task: the averaging set must be inferred from the same noisy observations whose variance it is intended to reduce.

The central difficulty is that repeated measurements of a latent point are unavailable. Every observation belongs to a pair, while the neighborhoods needed for averaging are themselves hidden. Moreover, a link that is informative only at short range cannot distinguish arbitrary large distances without a chaining principle. Thus distance estimation cannot begin until the data have first produced an observable surrogate for each latent point.

Our surrogate is a radius-controlled vertex cluster: a large set of sampled points in a small latent ball. Internal edges discover candidates, and disjoint external edges refine them and evaluate distances. Cluster-to-cluster averaging then separates stochastic fluctuation, controlled by cluster size, from geometric bias, controlled by cluster radius. The proof follows the chain internal similarity $\Rightarrow$ radius-controlled clusters $\Rightarrow$ link estimates $\Rightarrow$ latent distances.

Central question. From noisy and possibly missing pairwise observations, and without a representation of the latent space, what distance resolution is statistically attainable, and how much computation is required to attain it?

Contributions. An observable cluster replaces a hidden point when it is large enough for concentration and tight enough for bias control:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1. **Neighborhoods can be recovered without coordinates.** Noisy relations alone yield large, radius-controlled proxies for every sampled point under lower mass, link monotonicity, and the triangle inequality.
2. **Local averaging denoises the entire distance matrix.** Averaging between proxies separates noise from geometric bias, giving uniform recovery either directly or by chaining reliable short distances.
3. **Fixed accuracy is computationally inexpensive.** The expensive search stays inside a small working block, making the full procedure near-linear at fixed resolution; only shrinking resolution makes search dominant.
4. **Lower mass alone has an information limit.** Distinct folded metrics can generate nearly indistinguishable observations, sharpening the lower bound and exposing the unresolved statistical gap.

Why this is not a dimension substitution. Replacing r^d by a lower-mass profile does not remove manifold structure from a proof. Here occupancy supplies candidates, monotonicity orders them, pair averages validate them, and metric inequalities control bias; charts and tangent approximations never enter.

Paper organization. Section 2 gives the model and denoising principle; Section 3 states and interprets the guarantees; Section 4 explains the construction; Section 5 validates its error mechanism; Sections 6–7 discuss prior work and open questions. Complete proofs appear in the appendix.

2 Problem formulation and denoising principle

2.1 Observation model and assumptions

Let $(\mathcal{M}, d, \mu)$ be a metric probability space, identify $\mathcal{M}$ with $\operatorname{supp}(\mu)$, and assume $\operatorname{diam}(\mathcal{M}) < \infty$. Write $[n] := \{1, \dots, n\}$ for the vertex set. The latent points $X_1, \dots, X_n$ are sampled independently from μ, and for $U \subseteq [n]$ write $X_U := \{X_u : u \in U\}$. We denote the radius-r ball around x by $B(x, r)$. For each unordered pair $\{u, v\} \subseteq [n]$, we observe

$$Z_{uv} = B_{uv}\widetilde{Z}_{uv}, \qquad \mathbb{E}[\widetilde{Z}_{uv} \mid X_u, X_v] = p(d(X_u, X_v)), \tag{1}$$

where $B_{uv} \sim \operatorname{Bernoulli}(s)$ models missing observations and the centered noise in $\widetilde{Z}_{uv}$ is uniformly subgaussian. Conditional on the latent sample, the variables $\{Z_{uv} : u < v\}$ are independent and symmetric in their indices. This model contains weighted noisy measurements as well as Bernoulli similarity graphs. We state the paper for a non-increasing link p; a non-decreasing link is handled by changing signs.

Two local assumptions determine the available information. First, the sampling measure obeys the lower small-ball bound

$$\mu(B(x, r)) \geq \phi(r), \qquad x \in \operatorname{supp}(\mu), \quad 0 < r \leq r_\mu, \tag{2}$$

for a nondecreasing function ϕ. Second, for constants $0 < \ell_p \leq L_p < \infty$, the link is locally bi-Lipschitz:

$$\ell_p|a - b| \leq |p(a) - p(b)| \leq L_p|a - b|, \qquad a, b \in [0, r_p]. \tag{3}$$

The lower-mass profile controls how many observations can be averaged near an arbitrary latent point. The two link inequalities play different roles: the upper slope controls averaging bias, while the lower slope converts a link-value error back into a distance error.

The extraction algorithm knows s, ϕ, r_μ, r_p, ℓ_p, L_p, and $\sup_t |p(t)|$, but observes neither the points nor a representation of $\mathcal{M}$. The distance stage additionally uses p when inversion is required, and we write $\widehat{d}(v, w)$ for its estimate of $d(X_v, X_w)$. Constants below may depend only on these fixed model parameters.

2.2 Identifiability and geometric scope

The assumptions prevent three failures: (2) supplies local samples, a nonflat p separates distances, and fresh edges decouple certification from adaptive selection. Lower mass is a sampling rather than

smoothness condition; the upper link slope controls cluster bias, while the lower slope stabilizes inversion. Local denoising needs bi-Lipschitzness only on $[0, r_p]$, whereas long distances additionally require a global link or short-chain property.

The model includes regular manifolds with $\phi(r) \gtrsim r^d$ (Huang et al., 2024), but unlike classical reconstruction it assumes no reach, curvature, or ambient smoothness (Niyogi et al., 2008; Genovese et al., 2012; Aamari et al., 2019). It also covers metric trees, finite graphs, stratified spaces, rough subsets, and Ahlfors-regular fractals. On k discrete atoms it reduces to a stochastic block model with a lower bound on the smallest block mass (Holland et al., 1983; Abbe, 2018).

The scope is therefore intrinsic but deliberately limited. Bernoulli observations include soft random geometric graphs and latent-space networks (Penrose, 2003; Hoff et al., 2002), yet estimating an edge-probability matrix, a graphon, or a spectral position does not generally identify the generating metric (Athreya et al., 2018; Gao et al., 2015; Zhang et al., 2017; Issartel, 2021). Our guarantees concern d itself and consequently require an invertible relation between link values and distance on the range being estimated.

2.3 Observable neighborhood proxies

Here *observable neighborhood construction* means using only pairwise measurements to build a large vertex set whose latent points lie in a controlled ball. It does not estimate coordinates or locate points in an ambient space. The resulting cluster supplies repeated observations anchored near one hidden point: its size determines stochastic precision, while its radius determines the geometric averaging bias.

The definition below permits contamination because early candidates are only rough. The final extraction theorem uses exact clusters, but intermediate small-contamination statements are essential: they convert a global internal-average criterion into a progressively sharper local object without assuming that the true ball membership is observable.

Definition 2.1 (Radius-controlled cluster)**.** A nonempty $U \subseteq [n]$ is an (r, η)-cluster with center $x \in \mathcal{M}$ if $|\{u \in U : X_u \in B(x, r)\}| \geq (1 - \eta)|U|$. The case $\eta = 0$ is exact.

For $v \notin U$, define the normalized edge average and its conditional mean by

$$N_U(v) := \frac{1}{s|U|} \sum_{u \in U} Z_{uv}, \qquad S_U(X_v) := \frac{1}{|U|} \sum_{u \in U} p(d(X_u, X_v)). \tag{4}$$

Thus $\mathbb{E}[N_U(v) \mid X_U, X_v] = S_U(X_v)$, with the factor $1/s$ correcting for missing observations. If U lies near x, then $d(X_u, y)$ is close to $d(x, y)$ uniformly over $u \in U$. Whenever those distances remain in the informative link window, Lipschitzness transfers this radius control to link space. The statistic $N_U(v)$ is therefore a noisy, observable evaluation of the otherwise inaccessible function $y \mapsto p(d(x, y))$.

For two nonempty sets U, V, omitting diagonal pairs, let

$$N(U, V) := \frac{1}{s|D(U, V)|} \sum_{(u,v) \in D(U,V)} Z_{uv}, \qquad D(U, V) = \{(u, v) \in U \times V : u \neq v\}. \tag{5}$$

Write $p(U, V) := |D(U, V)|^{-1} \sum_{(u,v) \in D(U,V)} p(d(X_u, X_v))$ and $p(U) := p(U, U)$. When U and V represent centers x and y, respectively, the total error factors as

$$N(U, V) - p(d(x, y)) = \underbrace{N(U, V) - p(U, V)}_{\text{stochastic fluctuation}} + \underbrace{p(U, V) - p(d(x, y))}_{\text{cluster-radius bias}}. \tag{6}$$

Conditional concentration controls the first term; the triangle inequality and link regularity control the second. This identity is the contract between cluster extraction and latent-distance denoising: increasing the cluster size reduces variance, whereas decreasing its radius reduces bias.

2.4 Bias–variance balance and target scale

Fix a desired distance resolution $r > 0$. If $|U| = m$ and its evaluation edges are independent of its selection, subgaussian concentration gives $\max_{v \notin U} |N_U(v) - S_U(X_v)| \lesssim \log n / \sqrt{sm}$ with high

probability when $sm \gtrsim \log^2 n$. Since lower mass supplies $m \asymp n\phi(r)$ points in every radius-r neighborhood, resolving a link gap of order r requires

$$sn\phi(r)r^2 \gtrsim \log^2 n. \tag{7}$$

Thus (7) is a self-consistency condition: the radius determines how many points can be averaged, and that average must be accurate enough to resolve the same radius. The remaining challenge is selection rather than estimation—finding such averaging sets in the noisy graph while preserving the conditional independence used above.

The squared logarithm records the deliberately strong success level $1 - n^{-\omega(1)}$. For any fixed confidence exponent $A > 0$, retuning the concentration parameters gives success probability $1 - n^{-A}$ under the sharper requirement $sn\phi(r)r^2 \geq C_A \log n$, where C_A is a constant depending on A; the appendix audits this confidence-dependent form. Thus the polynomial part of the certified scale is set by $sn\phi(r)r^2$, while the displayed logarithm depends on the desired uniform confidence.

The two factors $\phi(r)$ and r^2 encode distinct costs. The first is geometric only through mass: a smaller ball supplies fewer potential replicates. The second is statistical: distinguishing two bounded or subgaussian means separated by order r requires order r^{-2} independent observations. Equation (7) is therefore not obtained by assigning an effective dimension; it follows by equating the resolution demanded from the link with the precision supplied by the local sample.

3 Guarantees, computational regimes, and limits

We first extract radius-controlled proxies, then convert them into distance estimates and compare the attainable scale with an information-theoretic obstruction. Full constants, admissible ranges, algorithms, and proofs are deferred to the appendix.

3.1 Upper bounds: from clusters to distances

Theorem 3.1 (Cluster extraction)**.** *Assume Equations* (2)–(3)*. There are constants $c, C > 0$ such that, whenever the target scale r satisfies* (7) *and lies inside the admissible local range, an algorithm takes the observed weighted graph as input and, with probability $1 - n^{-\omega(1)}$, returns for every target vertex v a set U_v satisfying*

$$v \in U_v, \qquad X_{U_v} \subseteq B(X_v, Cr), \qquad |U_v| \geq c\, n\phi(r).$$

The runtime is bounded by

$$\exp\left\{ C\, \frac{\log^2(1/\phi(r))}{sr^2} \right\} n \log^2 n,$$

and the number of distinct output clusters is at most the leading exponential factor.

Thus every sampled point has a large observable proxy wholly contained in an $O(r)$ ball: radius controls the bias in (6), and size controls its fluctuation.

The uniform guarantee lets one extracted family support the full distance matrix and accounts for the logarithmic scale factors. Exact containment prevents even a small set of distant members from creating nonvanishing bias outside the link's informative range.

Corollary 3.2 (Latent-distance denoising)**.** *Under the assumptions of Theorem 3.1:*

1. *If p is known and bi-Lipschitz on $[0, \operatorname{diam}(\mathcal{M})]$, all pairwise distances are recovered with uniform error $O(r)$.*
2. *If p is known but bi-Lipschitz only on $[0, r_p]$, distances within the local link window are recovered with error $O(r)$. If $\mathcal{M}$ is geodesic, shortest-path chaining of local estimates recovers all distances with the corresponding accumulated discretization error.*
3. *If p is unknown, the same construction combined with a multiscale midpoint-bisection step recovers the metric up to an unavoidable global dilation: for some unknown $D > 0$,*

$$\sup_{v,w} \left| D\widehat{d}(v, w) - d(X_v, X_w) \right| \leq O(r \log(1/r)).$$

These cases separate inversion from propagation: a global link is inverted directly, a local link is inverted on certified short pairs and propagated along chains, and an unknown link leaves an unavoidable global dilation. The same local cluster guarantee therefore remains meaningful on branching or disconnected spaces even when long distances cannot be chained.

Corollary 3.3 (Fixed-accuracy efficiency)**.** *If r, s, and $\phi(cr)$ are bounded below by positive constants, cluster extraction runs in $n \operatorname{polylog} n$ time. When the link is known and globally bi-Lipschitz, the subsequent distance-denoising stage has the same near-linear dependence on n.*

At fixed resolution, exhaustive search uses $n_1 \asymp \log\log n$ vertices and uniform refinement uses $n_2 \asymp \log^2 n$; the full sample is touched only after candidate reduction.

3.2 Interpreting resolution and runtime

Equation (7) is read directly from the lower-mass profile, without assigning a dimension to $\mathcal{M}$.

Ahlfors-regular spaces. If $\phi(r) \asymp r^d$ for an exponent $d > 0$, the smallest certified scale, denoted by r_{up}, is $r_{\mathrm{up}} \asymp (\log^2 n/(sn))^{1/(d+2)}$, or $n^{-1/(d+2)}$ up to logarithms when $s = 1$. The exponent combines local population r^d with the r^2 cost of resolving a mean gap of order r.

If the observation probability varies with n, write it as $s = s_n$. Sparsity enters through $s_n n\phi(r)$, so shrinking-scale denoising requires $s_n n/\log^2 n \to \infty$ and excludes bounded expected degree. Decreasing r reduces bias but shrinks the local population and enlarges the exhaustive search; fixed-accuracy efficiency is the constant-scale endpoint of this tradeoff. In a finite metric space, $\phi(r)$ is at least the smallest atom mass below the first nonzero distance, and clusters reduce to latent blocks.

The accuracy–runtime interaction is thus driven by the same quantity as statistical precision. At fixed r, every admissible ball has constant-order mass and the expensive search can be confined to a very small working set. When r shrinks, the algorithm must enlarge that set before it is likely to contain enough members of each hidden neighborhood, and enumeration rather than final averaging becomes the dominant cost.

Proposition 3.4 (Worst-case hardness of the exact candidate scan)**.** *Given a weighted graph, an integer m, and a threshold τ, deciding whether some m-vertex set U has internal average $N(U) \geq \tau$ is nondeterministic polynomial time (NP)-hard, even for complete observations with weights in $\{0, 1\}$.*

Indeed, for the adjacency matrix of an arbitrary graph and $\tau = 1$, such a set exists exactly when the graph contains an m-clique. Proposition 3.4 explains why exact enumeration is natural for the present objective, but it is only a worst-case statement: it neither rules out a different polynomial-time estimator under the metric-measure model nor establishes an average-case statistical–computational gap.

3.3 Information-theoretic limitation

To separate information from construction cost, we compare spaces with different latent-distance matrices but nearly indistinguishable observation laws.

For probability laws P and Q with P absolutely continuous with respect to Q, define their Kullback–Leibler (KL) divergence by $D_{\mathrm{KL}}(P\|Q) := \int \log(dP/dQ)\, dP$. We also define their total variation (TV) distance by $\mathrm{TV}(P, Q) := \sup_A |P(A) - Q(A)|$, where the supremum is over measurable events A.

Theorem 3.5 (Lower bound)**.** *For every nondecreasing lower-mass profile ϕ with $\phi(r) \to 0$ as $r \downarrow 0$, there exist lower-ϕ-regular metric probability spaces and monotone locally bi-Lipschitz links such that, for sequences $r_n \downarrow 0$ and $s_n \in (0, 1]$, uniform $o(r_n)$ latent-distance denoising is impossible whenever*

$$\sqrt{s_n}\, n\phi(r_n) r_n = o(1).$$

For the common profile $\phi(r) = r^d$ in the dense case, the upper and lower scales compare as follows.

The construction duplicates a small cap and folds its copy onto the original. Each cap contains about $n\phi(r)$ samples, so about $n^2\phi(r)^2$ cross-pairs are affected. A link displacement of order r contributes

Table 1: Accuracy scales when $\phi(r) = r^d$ and $s = 1$. Polylogarithmic factors are suppressed in the right column.

Result	Scale relation	Consequence for error
Upper bound	$nr^{d+2} \asymp \log^2 n$	$r \lesssim n^{-1/(d+2)} \operatorname{polylog} n$
Lower bound	$nr^{d+1} = o(1)$	$r \gtrsim n^{-1/(d+1)}$

$O(sr^2)$ Bernoulli KL divergence per affected pair, and Pinsker's inequality makes the aggregate TV scale $\sqrt{s}\, n\phi(r)r$. This second-order calculation is sharper than coupling the edges one at a time.

The two rows of Table 1 still leave an unresolved statistical gap under lower mass alone. A matching obstruction would need information scale $sn\phi(r)r^2$, rather than the quadratic cap-pair count used by folding; a centered hidden-partition construction is a natural candidate. Smooth models can instead exploit reach, curvature, and tangent regularity (Niyogi et al., 2008; Genovese et al., 2012; Aamari et al., 2019). Proposition 3.4 gives worst-case hardness of our exact scan, but we claim no average-case computational lower bound for data generated by the model.

Accordingly, the table should not be read as asserting optimality of the extraction rate. The upper row certifies what the cluster construction attains uniformly, while the lower row marks a resolution below which some admissible spaces cannot be distinguished. Closing their separation would require either a sharper estimator under the same assumptions or a stronger family of indistinguishable metric-measure models.

4 Algorithmic construction and proof logic

The proof separates discovery from certification: internal averages propose neighborhoods, cross-averages form a separated net, and fresh-edge scores refine it. Geometry converts population scores into radii; concentration transfers empirical scores to population values.

4.1 Pipeline and proof invariants

The sample and edge variables are split so each transition is independent of the data producing its input. The algorithm (i) enumerates near-maximal internal-average candidates on a small block, (ii) retains a separated covering family using disjoint cross-edges, (iii) reduces contamination on a fresh block, (iv) obtains uniform radius control on another, and (v) estimates distances only after fixing the clusters. Its invariant is rough coverage $\Rightarrow$ separated representatives $\Rightarrow$ small contamination $\Rightarrow$ uniform radius control. Thus only the first step is exhaustive, while every later certification remains conditionally independent.

Each arrow strengthens a different property. Coverage ensures that every sampled target is represented somewhere; separation prevents the candidate count from inheriting the combinatorial size of the initial search; contamination control turns an average geometric statement into a mostly local set; and uniform radius control finally makes every selected member safe for distance averaging. Keeping these properties distinct avoids using the desired conclusion—a tight neighborhood—inside the concentration argument that must establish it.

4.2 Candidate discovery by internal similarity

For $U \subseteq [n]$, let $N(U) := N(U, U)$. Fix a candidate scale $\rho > 0$ and a candidate size $m \asymp n\phi(\rho)$. Since p is non-increasing, $p(U) \le p(0)$, and near equality precludes many distant pairs. Conversely, each radius-ρ ball contains an m-subset with $p(U) \ge p(0) - O(\rho)$. Maximizing $N(U)$ over m-subsets therefore estimates $p(0)$ and finds neighborhood candidates simultaneously.

Let $\widehat{p}(0)$ denote the largest empirical internal average among the enumerated candidates. For a fixed constant $C > 0$, the search retains sets satisfying $N(U) \ge \widehat{p}(0) - C\rho$. Uniform concentration gives $p(0) - p(U) \le \Delta = O(\rho)$, which defines the internal deficit Δ. Decomposing this average yields an index $u_U \in U$ and a witness $\theta_U := X_{u_U}$ satisfying $|U|^{-1} \sum_{u \in U} [p(0) - p(d(X_u, \theta_U))] \le \Delta$. The inverse slope then gives $|U|^{-1} \sum_{u \in U} \min\{d(X_u, \theta_U), r_p\} \le \Delta/\ell_p$. For any comparison radius

$R \in (0, r_p]$, Markov's inequality converts this first-moment control into

$$\frac{|\{u \in U : X_u \notin B(\theta_U, R)\}|}{|U|} \le \frac{\Delta}{\ell_p R} \lesssim \frac{\rho}{R}, \qquad 0 < R \le r_p. \tag{8}$$

Thus internal deficit gives a witness, the inverse slope gives a truncated radial moment, and that moment gives tails at every informative radius. Probability only transfers empirical near-optimality and ensures that genuine balls occur among the candidates. Truncation at r_p is essential: a local link controls the fraction, but not the travel distance, of remaining outliers.

The witness is not intended as a coordinate estimate. It is used only to organize a deterministic inequality: averaging the pairwise deficit over the first endpoint guarantees that some endpoint has no larger one-sided deficit. This elementary averaging step replaces the unavailable center of a latent ball by an actual sampled point. The sliding bound in (8) then supplies all radii R simultaneously, which is what later refinements need when they trade the retained mass against score separation.

4.3 Candidate nets and fresh-edge purification

For a fresh vertex v, $N_U(v)$ concentrates around $S_U(X_v)$. Refinement needs no coordinate estimate, only a score gap around a witness x. Let Θ be the reference score and let $a_{\mathrm{in}} < a_{\mathrm{out}}$ be the inner and outer score deficits: inner points obey $S_U(y) \ge \Theta - a_{\mathrm{in}}$ and outer points obey $S_U(y) \le \Theta - a_{\mathrm{out}}$. Define the positive margin $\Gamma := a_{\mathrm{out}} - a_{\mathrm{in}}$. Thresholding midway is stable to noise below $\Gamma/4$. The first fresh block makes the exceptional fraction small; the second removes all exceptions uniformly over the seed net.

The midpoint threshold leaves a deterministic margin of $\Gamma/2$ on either side. Allocating half of that margin to empirical fluctuation and half to the imperfections of the rough seed makes the refinement composable: once a stage has certified the required score approximation, the next stage can condition on its selected set and invoke concentration only for previously unused edges. This is why one strong refinement cannot simply reuse the discovery edges.

Split the sample into a discovery block V_1, a first-refinement block V_2, and a final-refinement block V_3; primes denote smaller working subsets of their corresponding blocks. The search on V_1' gives rough coverage. Connecting candidates with cross-average near $\widehat{p}(0)$ and retaining a maximal independent set separates representatives while preserving coverage. Refinement on V_2' reduces contamination and refinement on V_3 removes it. Each transition uses disjoint edges, so conditional concentration is needed only over the surviving net; exhaustive search precedes the net, and the full sample is touched only after the family is small.

Maximality is the bridge between compression and coverage. Any discarded candidate is adjacent to a retained one and hence has a representative that is close in the empirical cross-average sense; the deterministic link comparison turns this adjacency into latent proximity. Nonadjacent retained candidates, conversely, occupy separated latent regions, so lower mass and total probability bound how many can survive. The net therefore reduces both runtime and the union bound used by the fresh-edge stages.

4.4 Two-scale bootstrap for a local link

With a global link, internal deficit controls the untruncated average radius and hence $|S_U(y) - p(d(\theta_U, y))| = O(\rho)$. With a local link, split U at $R \le r_p$: near points contribute at most $L_p R$, while (8) controls the fraction of far points. Inside the link window,

$$|S_U(y) - p(d(\theta_U, y))| \le L_p R + \frac{2\|p\|_\infty \Delta}{\ell_p R}. \tag{9}$$

Balancing at $R \asymp \sqrt{\Delta}$ gives $O(\sqrt{\rho})$ score error, so the unrestricted route searches at $\rho \asymp r^2$ for target scale r. A *link-safe* set has diameter below r_p, turning the truncated moment into an ordinary radius bound and restoring $O(\rho)$ bias. The observations construct a fuzzy certificate in three steps: run the unrestricted procedure at $r_0 \asymp \sqrt{r}$, use its clusters to certify common link-safe neighborhoods, and run the oracle-assisted procedure at r on fresh data. Because $r_0^2 \asymp r$, both rounds use the same local population and reduce to (7).

This square-root matching is the key algebraic reason that the bootstrap closes. Choosing a coarser first output than $\sqrt{r}$ would not certify the target neighborhoods sharply enough, whereas choosing a finer one would demand more local samples than the final scale provides. At $r_0 \asymp \sqrt{r}$, the first round pays the local-link loss exactly once; after link safety is established, the second round uses the linear $O(\rho)$ bias available inside the informative window.

Table 2: How link information determines the candidate scale. A link-safe certificate removes the square-root loss of the unrestricted local-link route.

Route	Candidate scale	Score error	Target error
Global bi-Lipschitz	$\rho \asymp r$	$O(\rho)$	$O(r)$
Local, no oracle	$\rho \asymp r^2$	$O(\sqrt{\rho})$	$O(r)$
Local, link-safe oracle	$\rho \asymp r$	$O(\rho)$	$O(r)$

4.5 Independence, working sets, and runtime

Let $V_1' \subseteq V_1$ be the exhaustive-search working set, with the smallest admissible size n_1 satisfying

$$sn_1\phi(\rho)\rho^2 \gtrsim \log^2 n_1, \qquad n_1 \geq \log\log n, \tag{10}$$

where $\rho \asymp r^2$ initially and $\rho \asymp r$ after link-safe certification. It enumerates $m_1 \asymp n_1\phi(\rho)$ subsets and passes only a maximal separated subfamily forward.

The first refinement uses V_2' of size n_2 satisfying

$$sn_2\phi(cr)r^2 \gtrsim \log^2 n. \tag{11}$$

This makes refinement uniform over V_3, which is accessed only after candidate reduction. At fixed dense resolution, $n_1 \asymp \log\log n$ and $n_2 \asymp \log^2 n$, giving $n \operatorname{polylog} n$ total time. At shrinking resolution, n_1 grows as $\phi(\rho)$ decreases, producing the bottleneck.

The two working-set equations serve different union bounds. Equation (10) only needs enough local witnesses for discovery within the small exhaustive block. Equation (11) must support a statement uniform over the final output vertices and therefore retains the ambient $\log^2 n$ factor. Separating these requirements is what prevents the combinatorial search from being performed on all n vertices.

4.6 Distance inversion and global propagation

Let r_{cl} denote the certified cluster radius, and let $\varepsilon_{\mathrm{pair}}$ denote the stochastic deviation of a cluster-pair average from its conditional mean. Suppose $X_{U_v} \subseteq B(X_v, r_{\mathrm{cl}})$ and $X_{U_w} \subseteq B(X_w, r_{\mathrm{cl}})$. Then every off-diagonal $(u, u') \in U_v \times U_w$ obeys $|d(X_u, X_{u'}) - d(X_v, X_w)| \leq 2r_{\mathrm{cl}}$. Inside the bi-Lipschitz range, averaging and concentration give

$$\begin{aligned} |N(U_v, U_w) - p(d(X_v, X_w))| \leq\ & |N(U_v, U_w) - p(U_v, U_w)| \\ & + |p(U_v, U_w) - p(d(X_v, X_w))| \leq \varepsilon_{\mathrm{pair}} + 2L_p r_{\mathrm{cl}}. \end{aligned} \tag{12}$$

The summands are stochastic error and radius bias. Inversion yields $|p^{-1}(N(U_v, U_w)) - d(X_v, X_w)| \leq (2L_p r_{\mathrm{cl}} + \varepsilon_{\mathrm{pair}})/\ell_p = O(r)$. Thus extraction need not recover a canonical ball: it needs enough points within a controlled target radius to balance both errors.

Notice that the two error terms are controlled by different outputs of Theorem 3.1. The size lower bound makes $\varepsilon_{\mathrm{pair}}$ small after a union bound over cluster pairs, while exact radius control bounds every summand in the population bias. This is also why merely recovering most members correctly is insufficient when unclassified outliers may lie beyond the link's informative range.

If p is only locally informative, certified estimates form a weighted graph. In a geodesic space, shortest paths chain these estimates globally, with local $O(r)$ error plus chain discretization error. Without such a property, large distances are not identified.

For unknown p, replacing d by Dd and composing p with the inverse dilation leaves the model unchanged; recovery up to one global factor is therefore the strongest identifiable conclusion.

5 QUANTITATIVE VALIDATION OF THE DENOISING MECHANISM

We test Equation (6) on $[0, 1]$ and an eight-arm metric tree. Distances are normalized to $[0, 1]$, $p(t) = 1 - t$, Gaussian noise has standard deviation 0.15, and edges are retained with $s = 0.75$. Retained edges are assigned to discovery, refinement, or evaluation with probabilities $0.375, 0.375, 0.25$, keeping evaluation independent of construction.

For each n, we use 24 targets and clusters of size $\lceil \sqrt{n} \rceil$. The *rough proxy* uses a discovery-block relation profile; the *refined proxy* combines it with an independent target-centered profile; and the *oracle proxy* uses latent nearest neighbors. All proxies invert disjoint cluster averages, whereas the *direct edge* baseline inverts one available evaluation edge. Table 3 reports diameter-normalized mean absolute error (MAE) over 20 repetitions.

Table 3: Normalized distance MAE over 20 repetitions. Proxies cover every target pair; the direct baseline covers 18.7% on average.

Space	n	Direct edge	Rough proxy	Refined proxy	Oracle proxy
Interval	300	$.107 \pm .012$	$.038 \pm .004$	$.031 \pm .004$	$.017 \pm .001$
	600	$.108 \pm .012$	$.027 \pm .003$	$.022 \pm .002$	$.012 \pm .001$
	1200	$.107 \pm .010$	$.018 \pm .002$	$.015 \pm .002$	$.008 \pm .001$
Metric tree	300	$.114 \pm .014$	$.083 \pm .015$	$.067 \pm .010$	$.046 \pm .007$
	600	$.117 \pm .015$	$.054 \pm .007$	$.041 \pm .005$	$.027 \pm .006$
	1200	$.112 \pm .011$	$.032 \pm .006$	$.025 \pm .005$	$.015 \pm .003$

Fresh refinement reduces rough-proxy MAE at every size: at $n = 1200$, from 0.018 to 0.015 on the interval and from 0.032 to 0.025 on the tree. The corresponding mean median cluster radii decrease from 0.102 to 0.088 and from 0.151 to 0.118, directly linking tighter neighborhoods to smaller bias. Refined MAE is then 86% and 77% below the direct baseline, covers every pair, and has 90th-percentile errors 0.031 and 0.053. Across the sensitivity runs in Appendix L, increasing s from 0.25 to 1 lowers refined MAE from 0.032 to 0.013 on the interval and from 0.082 to 0.024 on the tree; raising noise from 0.05 to 0.30 increases it from 0.005 to 0.034 and from 0.012 to 0.062. These controlled experiments diagnose the proof mechanism rather than establish a new rate.

6 RELATION TO PRIOR WORK

Metric-measure theory separates distance from sampling mass (Heinonen, 2001; Sturm, 2006); manifold and random-geometric-graph methods instead use smoothness, Euclidean structure, or hard thresholds (Niyogi et al., 2008; Genovese et al., 2012; Aamari et al., 2019; Huang et al., 2024; Araya Valdivia & De Castro, 2019; Arias-Castro et al., 2021). Graphon methods recover probability kernels rather than the generating metric (Gao et al., 2015; Zhang et al., 2017; Issartel, 2021). Our lower-mass analysis directly targets intrinsic distances.

7 OPEN QUESTIONS

1. Can the method adapt to unknown ϕ and link parameters while preserving the same uniform scale condition?
2. Is fine-scale metric-measure denoising average-case hard, or can a different polynomial-time algorithm avoid Proposition 3.4?
3. Can a centered hidden-partition construction match the scales in Table 1?
4. Which weak chain, doubling, or spectral assumptions sharpen the rate without restoring a smooth manifold model?

AI USE STATEMENT

We used generative AI tools solely for language editing and proofreading. We did not use generative AI tools to generate synthetic datasets; develop theoretical or conceptual models; formulate mathematical claims; contribute to proofs; propose hypotheses; design the methodology or experiments; implement the proposed methods; process or analyze the data; or interpret the experimental results. All claims, numerical results, scripts, and tables were produced and verified by the authors using

the original source files and experimental outputs. We reviewed all AI-assisted text and take full responsibility for the final content of this work.

Reproducibility statement

The information and materials needed to reproduce our experiments are provided in the appendix and the anonymized supplementary material. The experiments use open-weight instruction backbones, fixed random seeds, public benchmarks, and saved JSONL traces containing full and single-component counterfactual interventions. The supplementary repository records the trace schema, generation commands, calibration scripts, policy summaries, and resource footprints. We also provide anonymized source code, trace-processing scripts, task identifiers, and the scripts used to generate the reported summary tables.

References

Eddie Aamari, Jisu Kim, Frédéric Chazal, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Estimating the reach of a manifold. *Electronic Journal of Statistics*, 13(1):1359–1399, 2019. doi: 10.1214/19-EJS1551.

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

Ernesto Araya Valdivia and Yohann De Castro. Latent distance estimation for random geometric graphs. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Ery Arias-Castro, Antoine Channarond, Bruno Pelletier, and Nicolas Verzelen. On the estimation of latent distances using graph distances. *Electronic Journal of Statistics*, 15(1):722–747, 2021.

Avanti Athreya, Donniell E. Fishkind, Minh Tang, Carey E. Priebe, Youngser Park, Joshua T. Vogelstein, Keith Levin, Vince Lyzinski, Yichen Qin, and Daniel L. Sussman. Statistical inference on random dot product graphs: A survey. *Journal of Machine Learning Research*, 18(226):1–92, 2018.

Chao Gao, Yu Lu, and Harrison H. Zhou. Rate-optimal graphon estimation. *The Annals of Statistics*, 43(6):2624–2652, 2015. doi: 10.1214/15-AOS1354.

Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *Journal of Machine Learning Research*, 13(43):1263–1291, 2012.

Juha Heinonen. *Lectures on Analysis on Metric Spaces*. Springer, New York, 2001.

Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

Han Huang, Pakawut Jiradilok, and Elchanan Mossel. Reconstructing the geometry of random geometric graphs. In *Proceedings of the 37th Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 2519–2521, 2024.

Yann Issartel. On the estimation of network complexity: Dimension of graphons. *Journal of Machine Learning Research*, 22(191):1–62, 2021.

Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1–3):419–441, 2008. doi: 10.1007/s00454-008-9053-2.

Mathew Penrose. *Random Geometric Graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, 2003.

Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta Mathematica*, 196(1):65–131, 2006. doi: 10.1007/s11511-006-0002-8.

Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2009. doi: 10.1007/978-3-540-71050-9.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. Estimating network edge probabilities by neighborhood smoothing. *Biometrika*, 104(4):771–783, 2017. doi: 10.1093/biomet/asx042.

A MODEL, NOTATION, AND PROOF ORGANIZATION

Appendix organization. The appendix is organized as follows.

- **Model, notation, and basic events:** fixes the metric-measure model, the sparse observation model, local link regularity, radius-controlled clusters, and the high-probability events used in every proof.
- **Probability and occupancy tools:** proves the uniform small-ball occupancy, navigation concentration, and empirical pair-average bounds needed by the construction.
- **Internal-average candidate construction:** shows that maximizing an observable internal average estimates $p(0)$ and produces candidates with a latent radius witness.
- **Geometry of rough neighborhood candidates:** derives truncated radial moments, local score approximations, link-admissible cross-average bounds, and the square-root loss caused by a merely local link.
- **Candidate nets from observable families:** turns the exponentially large candidate family into a separated covering family, both with and without a link-safe certificate.
- **Fresh-edge cluster refinement:** proves the thresholding lemma that converts a rough candidate first into a low-contamination cluster and then into an exact radius-controlled cluster.
- **Three-block neighborhood extraction:** combines discovery, net construction, and two independent refinements, including the working-set and runtime analysis.
- **Link-safe bootstrap:** constructs a fuzzy link-safe certificate from a coarse run and uses it in a fine run to prove Theorem 3.1 at the target scale.
- **End-to-end invariant and parameter audit:** records the invariant carried by the algorithm and verifies scale matching, entropy, failure probabilities, and admissible radii directly from the master condition.
- **From clusters to distances:** proves the global-link, local-link, chaining, and fixed-accuracy conclusions in Corollaries 3.2–3.3.
- **Folding lower bound:** proves Theorem 3.5 by a lifted KL comparison between a space and a folded copy whose observations are asymptotically indistinguishable.
- **Supplementary experiments:** records the rough-versus-refined radius ablation, observation-rate and noise sensitivity, and measured runtime behind Section 5.

The ordering follows logical dependence. In particular, each probabilistic statement appears before its first algorithmic use, and the proof of Theorem 3.1 becomes a forward chain rather than a collection of deferred arguments.

There are two complementary ways to read the appendix. A correctness-first reading follows the sections in order: probability events validate internal candidates; candidate geometry produces witnesses; the observable graph selects a net; two fresh-edge passes purify that net; and the coarse-to-fine bootstrap closes the local-link argument. A parameter-first reading begins with Appendix I, which lists the invariant and the scale budget, and then follows its citations backward to the lemma supplying each transition. The distance and lower-bound sections are logically independent once cluster extraction has been established. The former treats exact clusters as repeated measurements of hidden points, whereas the latter constructs two spaces whose observations agree even though a sampled distance changes. This organization separates the constructive upper bound from the information-theoretic obstruction while making their dependence on the same lower-mass profile explicit.

Throughout, statements described as deterministic are conditioned on the latent sample and on the explicitly named good events. They contain no further probability. Statements described as observable use only edge variables and fixed model parameters, even when their proofs introduce latent witnesses. This convention distinguishes an implementable construction from the latent objects used solely to certify it.

How the proof closes. It is useful to distinguish the three kinds of error that recur below. An *empirical error* is the deviation of an observed average from its conditional mean; it is controlled by

fresh-edge concentration. A *geometric bias* is the difference between that conditional mean and the link value at a latent center; it is controlled by the radius or truncated radial moment of the averaging set. An *identification error* arises when a link-value error is converted into distance and costs the reciprocal slope $1/\ell_p$. The argument never merges these quantities into a single black-box estimate. Candidate discovery first controls a population internal deficit, the rough-geometry lemmas convert it into a spatial statement, and refinement then spends an independent empirical budget to make that spatial statement uniform. The bootstrap is needed only to certify that the inverse-slope step remains inside the local link window. Finally, the distance section reuses the same decomposition once more, now with two exact clusters instead of one rough candidate. This separation makes clear which conclusions use probability, which use only the triangle inequality, and which genuinely require local identifiability of the link.

A.1 Model and notation

We work conditionally on the latent sample whenever concentration of edge statistics is invoked. Let V be a finite vertex set with $|V| = n$, and sample the variables $(X_v)_{v\in V}$ independently from μ. For $u < v$, write

$$Z_{uv} = B_{uv}\widetilde{Z}_{uv}, \qquad B_{uv} \sim \text{Bernoulli}(s), \qquad \mathbb{E}[\widetilde{Z}_{uv} \mid X_u, X_v] = p(d(X_u, X_v)).$$

The centered variables $\widetilde{Z}_{uv} - p(d(X_u, X_v))$ are uniformly K_{sg}-subgaussian, and all unordered-pair variables are conditionally independent. Put

$$M_p := \sup_{0\le t\le \text{diam}(\mathcal{M})} |p(t)|.$$

The constants below may depend on $K_{\text{sg}}, M_p, L_p, \ell_p, r_p, r_\mu$, but never on n, s, or the target scale.

The measure assumption is used only through

$$\mu(B(x, r)) \ge \phi(r), \qquad x \in \mathcal{M}, \quad 0 < r \le r_\mu, \tag{13}$$

where ϕ is nonnegative and nondecreasing. The link is non-increasing and satisfies

$$\ell_p|a - b| \le |p(a) - p(b)| \le L_p|a - b|, \qquad a, b \in [0, r_p]. \tag{14}$$

The upper inequality controls the bias created by averaging over a cluster; the lower inequality turns an observable score gap back into a metric separation. Neither inequality is invoked outside $[0, r_p]$ unless global regularity is explicitly assumed.

For nonempty $U \subseteq V$ and $x \in \mathcal{M}$, define the average radius

$$\text{rad}_{\text{av}}(U, x) := \frac{1}{|U|}\sum_{u\in U} d(X_u, x).$$

The score S_U, the normalized navigation average $N_U(v)$, the pair average $N(U, W)$, and its conditional mean $p(U, W)$ are those in Equations (4)–(5). A set U is *link-admissible* if $\text{diam}(X_U) \le r_p$, and is λ-*link-safe* if $\text{diam}(X_U) \le \lambda r_p$ for a fixed $\lambda < 1/4$.

For constants $0 < \alpha < \lambda < 1/4$, a map

$$\mathcal{O}_{\text{safe}} : V \times V \longrightarrow \{0, 1\}$$

is an (α, λ)-*fuzzy link-safe certificate* if

$$d(X_u, X_v) \le \alpha r_p \Longrightarrow \mathcal{O}_{\text{safe}}(u, v) = 1, \qquad d(X_u, X_v) > \lambda r_p \Longrightarrow \mathcal{O}_{\text{safe}}(u, v) = 0.$$

No condition is imposed in the intermediate gray zone. This is an analytical interface constructed from fresh observations in Appendix H; it is not side information available to the algorithm.

We use two events. For $W \subseteq V$ and $r > 0$, let

$$\mathcal{E}_{\text{occ}}(W, r) := \left\{ |\{v \in W : X_v \in B(x, r)\}| \ge \frac{|W|\phi(r/3)}{2} \text{ for every } x \in \mathcal{M} \right\}. \tag{15}$$

For $m \ge 2$ and $\lambda > 0$, let

$$\mathcal{E}_{\text{avg}}(W, \lambda, m) := \{|N(U_1, U_2) - p(U_1, U_2)| \le \lambda \text{ for all } U_1, U_2 \subseteq W,\ |U_i| \ge m\}. \tag{16}$$

The diagonal is omitted in both averages, so overlapping sets cause no ambiguity.

For navigation from a fixed candidate into a fresh block, write

$$\mathcal{E}_{\mathrm{nav}}(U, W; t) := \left\{ \max_{v \in W} |N_U(v) - S_U(X_v)| \le t \right\}. \tag{17}$$

The three events (15), (16), and (17) separate the proof into population geometry and conditional concentration. Once they hold, all remaining implications in the extraction argument are deterministic.

A.2 Deterministic interfaces used by the proof

The proof repeatedly passes through the same three interfaces. We record them here so that later arguments can cite the relevant input rather than silently changing the meaning of a cluster or a score.

Definition A.1 (Deficit, radial profile, and score error)**.** For a nonempty vertex set U and a witness $x \in \mathcal{M}$, define

$$\Delta_{\mathrm{int}}(U) := p(0) - p(U), \qquad \tau_U(x; R) := \frac{|\{u \in U : d(X_u, x) > R\}|}{|U|},$$

and, for a query point y for which $p(d(x, y))$ is meaningful in the local window,

$$\beta_U(x; y) := |S_U(y) - p(d(x, y))|.$$

The first quantity is an internal link deficit, the second is a deterministic tail profile, and the third is the geometric averaging bias of the navigation score.

An internal-average argument first proves that $\Delta_{\mathrm{int}}(U)$ is small. Lemma C.3 then selects $x = \theta_U$ and controls $\tau_U(\theta_U; R)$ simultaneously for every $R \le r_p$. Lemma D.1 finally converts that profile into a bound for β_U. These arrows are one-way: a small empirical score at a single query point does not itself imply that U is geometrically concentrated, and a small tail fraction does not identify a unique center. The construction needs only existence of a witness, not uniqueness.

Lemma A.2 (Three-term error composition)**.** *Let U be fixed before its navigation edges to v are revealed, and let $x \in \mathcal{M}$. On $\mathcal{E}_{\mathrm{nav}}(U, \{v\}; t)$,*

$$|N_U(v) - p(d(x, X_v))| \le t + \beta_U(x; X_v). \tag{18}$$

If p is invertible on an interval containing $d(x, X_v)$ and the right-hand side is ε, then projection of $N_U(v)$ to the image of that interval followed by inversion has distance error at most ε/ℓ_p.

Proof. Insert the conditional mean $S_U(X_v)$ and apply the triangle inequality:

$$|N_U(v) - p(d(x, X_v))| \le |N_U(v) - S_U(X_v)| + |S_U(X_v) - p(d(x, X_v))|.$$

The two terms are bounded by t and $\beta_U(x; X_v)$, respectively. Metric projection onto an interval cannot increase absolute error. The inverse local slope in (14) then gives the last assertion. □

The same decomposition will later be applied to a pair of clusters. There the empirical term is $|N(U_1, U_2) - p(U_1, U_2)|$, the geometric term is $|p(U_1, U_2) - p(d(x_1, x_2))|$, and inversion again costs $1/\ell_p$. Keeping the three terms separate matters because they scale differently: empirical fluctuation is controlled by s and cluster size, geometric bias by cluster radius, and inversion only by link identifiability.

Lemma A.3 (Admissible constant hierarchy)**.** *All fixed constants used below can be chosen in the following order:*

$$\alpha, \lambda; \quad K_{\mathrm{sep}}; \quad K_{\mathrm{cov}}; \quad K_2, K_3; \quad c_\rho; \quad c_{\mathrm{scale}}; \quad C_{\mathrm{prob}}.$$

Here $0 < \alpha < \lambda < 1/4$ depend only on L_p/ℓ_p; the K's are sufficiently large constants; c_ρ and c_{scale} are then sufficiently small; and the final probability constant is sufficiently large. With this order, every radius appearing in a local-link comparison is at most r_p and every radius used in occupancy is at most r_μ.

Proof. Choose the oracle margins first, since they determine the fixed fraction of the link window available to later comparisons. The separation constants are then chosen to dominate the ratios L_p/ℓ_p and M_p/ℓ_p in the score-gap arguments. Once these multiplicative losses are fixed, choose the candidate-scale constant so that the largest resulting radius is below the oracle margin. The admissible target-scale constant can then enforce all restrictions relative to r_p and r_μ. Finally enlarge the concentration constant; this affects only sample-size inequalities and does not change any deterministic radius comparison. No circular choice remains. □

This hierarchy explains the convention that symbols c and C may change from line to line. Every such change can be realized by decreasing the last unused small constant or increasing the last unused large constant. None depends on n, s, r, or the particular metric space beyond the fixed model parameters listed above.

B Probability and occupancy tools

B.1 Uniform occupancy

Lemma B.1 (Net cardinality). *If $(\mathcal{M}, d, \mu)$ is lower-ϕ-regular and $\mathcal{N}$ is a maximal δ-separated subset of $\mathcal{M}$, then $\mathcal{N}$ is a δ-net and*

$$|\mathcal{N}| \le \phi(\delta/2)^{-1}.$$

Proof. Maximality gives the net property. The balls $B(x, \delta/2)$, $x \in \mathcal{N}$, are pairwise disjoint, hence

$$1 = \mu(\mathcal{M}) \ge \sum_{x \in \mathcal{N}} \mu(B(x, \delta/2)) \ge |\mathcal{N}|\phi(\delta/2).$$

□

Lemma B.2 (Uniform lower occupancy). *Let $n_\star = |W|$. If $r \le r_\mu$ and*

$$n_\star \phi(r/6) \ge C \log^2 n_\star,$$

then $\mathcal{E}_{\mathrm{occ}}(W, r)$ holds with probability $1 - n_\star^{-\omega(1)}$.

Proof. Let $\mathcal{N}$ be a maximal $r/3$-separated set. By Lemma B.1, $|\mathcal{N}| \le 1/\phi(r/6)$. For each $z \in \mathcal{N}$, the number of samples in $B(z, r/3)$ is binomial with mean at least $n_\star\phi(r/3)$. The lower Chernoff bound and a union bound give

$$\mathbb{P}\big(\exists z \in \mathcal{N} : |W \cap B(z, r/3)| < \tfrac{1}{2} n_\star \phi(r/3)\big) \le \frac{1}{\phi(r/6)} \exp\{-c n_\star \phi(r/3)\} = n_\star^{-\omega(1)}.$$

For any $x \in \mathcal{M}$, select $z \in \mathcal{N}$ with $d(x, z) \le r/3$. Then $B(z, r/3) \subseteq B(x, r)$, which proves (15). □

Corollary B.3 (Simultaneous occupancy across a block design). *Let $W_1, \dots, W_J$ be disjoint vertex blocks, where J is fixed and $|W_j| \ge c_0 n$. Let $\mathcal{R}$ be a fixed finite family of radii with $|\mathcal{R}| \le n^A$. If, for every $r \in \mathcal{R}$,*

$$n\phi(r/6) \ge C \log^2 n,$$

then with probability $1 - n^{-\omega(1)}$, $\mathcal{E}_{\mathrm{occ}}(W_j, r)$ holds for every $j \le J$ and every $r \in \mathcal{R}$.

Proof. Apply Lemma B.2 to each pair (W_j, r). Replacing n by $|W_j|$ loses only a fixed factor in the exponent because $|W_j| \ge c_0 n$. The failure probability for one pair is smaller than n^{-B} for every fixed B once the leading constant is large. A union bound over $J|\mathcal{R}| \le Jn^A$ pairs preserves superpolynomial success. □

Lemma B.4 (Occupancy implies a sampled net). *If $\mathcal{E}_{\mathrm{occ}}(W, r)$ holds and $|W|\phi(r/3) \ge 2$, then X_W is an r-net of $\mathcal{M}$. More generally, every $x \in \mathcal{M}$ has at least $|W|\phi(r/3)/2$ sampled representatives within distance r.*

Proof. The second assertion is exactly the event definition (15). Its lower bound is at least one under the numerical assumption, so every radius-r ball contains a sampled point. This is the net property. □

The multiplicity statement, rather than mere net coverage, is the part used by the algorithm. A single representative would locate a neighborhood only in an analysis that already knew latent positions. Here a population of order $n\phi(r)$ is needed because the neighborhood must itself serve as an averaging device. This is why the same lower-mass term appears both in the existence of candidates and in their eventual concentration rate.

B.2 Conditional concentration

Lemma B.5 (Navigation concentration)**.** *For fixed, disjoint $U, W \subseteq V$, $m = |U|$, and $t \in (0, 1]$,*

$$\mathbb{P}(|N_U(v) - S_U(X_v)| > t \mid X_U, X_W) \le C \exp(-csmt^2)$$

for each $v \in W$. Consequently, if $sm \ge C \log^2 n_\star$ and $|W| \le n_\star^A$, then

$$\max_{v \in W} |N_U(v) - S_U(X_v)| \le C \frac{\log n_\star}{\sqrt{sm}}$$

with conditional probability $1 - n_\star^{-\omega(1)}$.

Proof. Fix the latent points and write $a_u = p(d(X_u, X_v))$ and $\xi_u = \widetilde{Z}_{uv} - a_u$. Then

$$N_U(v) - S_U(X_v) = \frac{1}{sm} \sum_{u \in U} (B_{uv} - s) a_u + \frac{1}{sm} \sum_{u \in U} B_{uv} \xi_u =: T_1 + T_2.$$

The first sum consists of independent centered variables bounded by M_p, with total variance at most smM_p^2; Bernstein's inequality gives $\mathbb{P}(|T_1| > t/2) \le 2 \exp(-csm \min\{t^2, 1\})$. Let $N_B = \sum_{u \in U} B_{uv}$. Chernoff's inequality gives $\mathbb{P}(N_B > 2sm) \le e^{-csm}$. Conditional on $N_B \le 2sm$ and on the Bernoulli variables, the numerator of T_2 is subgaussian with parameter at most $K_{\mathrm{sg}} \sqrt{2sm}$. Hence $\mathbb{P}(|T_2| > t/2) \le 3 \exp(-csm \min\{t^2, 1\})$. Combining the bounds proves the first assertion. Set $t = C \log n_\star / \sqrt{sm}$ and union-bound over W for the second. □

Lemma B.6 (Concentration for a conditionally selected family)**.** *Let $\mathcal{F}$ be a sigma-field containing the latent sample and all observations used to construct a finite family $\mathcal{A}$ of vertex sets. Suppose $\mathcal{A}$ is $\mathcal{F}$-measurable, $|\mathcal{A}| \le K$ almost surely, every $U \in \mathcal{A}$ has $|U| \ge m$, and the edges from U to a fresh block W are independent of $\mathcal{F}$. Then, for $t \in (0, 1]$,*

$$\Pr\left\{ \max_{U \in \mathcal{A}} \max_{v \in W} |N_U(v) - S_U(X_v)| > t \right\} \le CK|W| e^{-csmt^2}. \tag{19}$$

In particular, the maximum is at most

$$C \sqrt{\frac{\log(K|W|) + u}{sm}}$$

with probability at least $1 - e^{-u}$, whenever the displayed quantity is at most one.

Proof. Condition on $\mathcal{F}$. The family $\mathcal{A}$, its cardinality, and all latent points are then fixed, while every navigation statistic uses unrevealed independent edges. Lemma B.5 applies to each ordered pair (U, v). A conditional union bound gives the right-hand side of (19). The bound no longer depends on the realized family, so integrating over $\mathcal{F}$ proves the unconditional statement. Solve $CK|W|e^{-csmt^2} \le e^{-u}$ for t to obtain the last display. □

Corollary B.7 (Uniform score accuracy after net reduction)**.** *Suppose a discovery stage outputs a family $\mathcal{N}$ with $|\mathcal{N}| \le n_1$ and all seeds have size at least m. On a fresh block W of size at most n, if*

$$sm\Gamma^2 \ge C(\log n + \log n_1),$$

then all navigation scores are within $\Gamma/4$ of their population values with probability $1 - n^{-\omega(1)}$ after increasing the right-hand side to $C \log^2 n$.

Proof. Use Lemma B.6 with $K = n_1$, $t = \Gamma/4$, and $u = \log^2 n$. Since $n_1 \le n$ in the applications, the logarithmic entropy of the surviving family is negligible relative to u. □

This corollary formalizes the role of the maximal independent set. Before reduction, the number of candidate subsets may be exponential in n_1, and a navigation union bound over that family would spend the same entropy a second time. After reduction, the stochastic cost is only $\log n_1 + \log n$. The exhaustive search is therefore paid for once, on internal edges of the smallest block, rather than at every later refinement.

Corollary B.8 (Exceptional-vertex count)**.** *Under the hypotheses of Lemma B.5, fix a gap* $\Gamma \in (0, 1]$ *and define*

$$I_\Gamma(U, W) := \{v \in W : |N_U(v) - S_U(X_v)| > \Gamma/4\}.$$

Conditionally on X_U, X_W*, the indicators defining* I_Γ *are independent and each has mean at most*

$$q(m, \Gamma) := C \exp(-csm\Gamma^2).$$

Consequently, for every reference size $N \ge |W|$*,*

$$|I_\Gamma(U, W)| \le 2q(m, \Gamma)|W| + 2\log^2 N \tag{20}$$

with conditional probability at least $1 - e^{-c \log^2 N}$*.*

Proof. The independence follows because different $v \in W$ use disjoint unordered pairs $\{u, v\}$, $u \in U$. Lemma B.5 gives the bound on each mean. Conditional stochastic domination by $\mathrm{Binomial}(|W|, q)$ and the upper Chernoff inequality give

$$\mathbb{P}\big(|I_\Gamma| > 2q|W| + 2\log^2 N \mid X_U, X_W\big) \le e^{-c \log^2 N}.$$

The deliberately loose logarithmic term is useful because the bound will later be unioned over a data-dependent but separated seed family. □

Lemma B.9 (Uniform pair-average concentration)**.** *Let* $n_\star = |W|$*,* $\varphi \in (0, 1)$*, and* $m = \lceil \varphi n_\star \rceil \ge 2$*. If*

$$sn_\star \varphi \lambda^2 \ge C \log(e/\varphi),$$

then, conditionally on X_W*,*

$$\mathbb{P}\big(\mathcal{E}_{\mathrm{avg}}(W, \lambda, m)^c \mid X_W\big) \le \exp\{-c\varphi n_\star \log(e/\varphi)\}.$$

Proof. Fix $U_1, U_2 \subseteq W$ with sizes $a, b \ge m$. For an unordered pair $e = \{u, v\}$, let $w_e \in \{0, 1, 2\}$ be its multiplicity in the ordered off-diagonal set $D(U_1, U_2)$. With $D = |D(U_1, U_2)| = \sum_e w_e$, the same decomposition as in Lemma B.5 yields

$$N(U_1, U_2) - p(U_1, U_2) = \frac{1}{sD} \sum_e w_e (B_e - s) a_e + \frac{1}{sD} \sum_e w_e B_e \xi_e.$$

Because $w_e \le 2$, $\sum_e w_e^2 \le 2D$, and $D \ge ab/2$, Bernstein and subgaussian concentration give

$$\mathbb{P}\big(|N(U_1, U_2) - p(U_1, U_2)| > \lambda \mid X_W\big) \le 4e^{-csD\lambda^2} \le 4e^{-csab\lambda^2/2}.$$

For $a = \alpha n_\star$ and $b = \beta n_\star$, the number of pairs of sets is at most

$$\binom{n_\star}{a}\binom{n_\star}{b} \le \exp\{(\alpha + \beta) n_\star \log(e/\varphi)\}.$$

Since $\alpha, \beta \ge \varphi$ and $\alpha + \beta \le 2\alpha\beta/\varphi$, the assumed scale condition makes the concentration exponent dominate this entropy. Summing over at most $n_\star^2$ cardinality pairs gives the stated bound. □

B.3 Detailed entropy calculation for pair averages

The preceding lemma is the only place where the proof pays for all possible subsets rather than for a selected family. We expand the calculation because the normalization, overlapping candidates, and missing observations otherwise obscure why the effective exponent is linear in the smaller set size rather than quadratic in the working-set size.

Lemma B.10 (Fixed-cardinality empirical-process bound)**.** *Fix integers $2 \le a, b \le n_\star$. Conditional on X_W, for every $\lambda \in (0, 1]$,*

$$\Pr\left\{ \max_{\substack{|U_1|=a \\ |U_2|=b}} |N(U_1, U_2) - p(U_1, U_2)| > \lambda \mid X_W \right\} \le 4 \exp\left\{ a \log \frac{en_\star}{a} + b \log \frac{en_\star}{b} - csab\lambda^2 \right\}. \tag{21}$$

The bound remains valid when U_1 and U_2 overlap.

Proof. For fixed U_1, U_2, write the statistic as a weighted sum over unordered edges, as in Lemma B.9. An edge in the overlap can occur twice in the ordered cross-product, but its weight is never larger than two. Hence

$$\sum_e w_e^2 \le 2 \sum_e w_e = 2D, \qquad D = ab - |U_1 \cap U_2| \ge ab - \min\{a, b\} \ge ab/2.$$

The missingness term is a sum of bounded centered variables whose total conditional variance is at most CsD. The noise term, conditional on the Bernoulli selectors, is subgaussian with squared scale at most CsD except on an event of probability e^{-csD}. Bernstein and subgaussian tails therefore give

$$\Pr\{|N(U_1, U_2) - p(U_1, U_2)| > \lambda \mid X_W\} \le 4e^{-csab\lambda^2}.$$

There are $\binom{n_\star}{a}\binom{n_\star}{b}$ ordered choices of the two sets. Apply a union bound and $\binom{n_\star}{a} \le (en_\star/a)^a$ to obtain (21). The multiplicity calculation already accounts for overlap, so disjointness is unnecessary. □

Lemma B.11 (Summation over all large cardinalities)**.** *Let $m = \lceil \varphi n_\star \rceil$ and suppose*

$$sn_\star \varphi \lambda^2 \ge C \log(e/\varphi). \tag{22}$$

Then the sum of the right-hand side of (21) *over every $a, b \ge m$ is at most*

$$\exp\{-c\varphi n_\star \log(e/\varphi)\}. \tag{23}$$

Proof. Write $a = \alpha n_\star$ and $b = \beta n_\star$, where $\alpha, \beta \ge \varphi$. The entropy exponent is bounded by

$$n_\star \left[\alpha \log(e/\alpha) + \beta \log(e/\beta)\right] \le n_\star(\alpha + \beta) \log(e/\varphi).$$

Since $\alpha, \beta \ge \varphi$, $\alpha + \beta \le 2\alpha\beta/\varphi$. Under (22), the concentration term satisfies

$$csab\lambda^2 = cs\alpha\beta n_\star^2 \lambda^2 \ge C' \frac{\alpha\beta}{\varphi} n_\star \log(e/\varphi),$$

which dominates the entropy by a fixed factor. At the smallest cardinalities, the remaining negative exponent is at least $c\varphi n_\star \log(e/\varphi)$. There are at most $n_\star^2$ cardinality pairs. Enlarging the constant in (22) absorbs this polynomial factor and proves (23). □

Two features are worth noting. First, the event is uniform over overlapping and data-independent subsets because it is proved before any candidate is selected. Second, the scale condition contains $sn_\star\varphi\lambda^2$, not $sn_\star^2\varphi^2\lambda^2$. A fixed pair average enjoys the latter quadratic number of contributing edges, but the class of all subsets has entropy of order $n_\star\varphi\log(e/\varphi)$. Dividing the concentration exponent by this entropy leaves exactly the effective local sample size $sn_\star\varphi$ that appears throughout the theorem.

Lemma B.12 (Effect of overlap and diagonal removal)**.** *For nonempty $U_1, U_2 \subseteq W$, let*

$$D(U_1, U_2) := \{(u, v) \in U_1 \times U_2 : u \neq v\}.$$

Then

$$|D(U_1, U_2)| = |U_1||U_2| - |U_1 \cap U_2|. \tag{24}$$

If $|U_1|, |U_2| \geq m \geq 2$, then

$$\frac{1}{2}|U_1||U_2| \leq |D(U_1, U_2)| \leq |U_1||U_2|. \tag{25}$$

Moreover, for arbitrary numbers c_u,

$$\frac{1}{|D(U_1, U_2)|} \sum_{(u,v)\in D(U_1,U_2)} c_u \leq 2\frac{1}{|U_1|} \sum_{u\in U_1} |c_u| \tag{26}$$

after absolute values are applied to the left-hand summands. The symmetric statement holds for quantities depending only on v.

Proof. The full Cartesian product has $|U_1||U_2|$ ordered pairs. A diagonal pair (u, u) occurs exactly when $u \in U_1 \cap U_2$, proving (24). Since $|U_1 \cap U_2| \leq \min\{|U_1|, |U_2|\} \leq |U_1||U_2|/2$ for both sizes at least two, (25) follows.

For fixed $u \in U_1$, the number of allowed second endpoints is at most $|U_2|$. Therefore the absolute numerator in (26) is at most $|U_2| \sum_{u\in U_1} |c_u|$. Divide by the lower bound $|D(U_1, U_2)| \geq |U_1||U_2|/2$ to obtain the result. □

This bookkeeping justifies the constants in every cross-average bias bound. Candidate sets may overlap heavily, and internal averages correspond to the extreme case $U_1 = U_2$. Removing the diagonal avoids undefined self-observations but changes normalization by at most a factor two. Thus endpoint contributions can still be controlled by ordinary averages over each set, with no disjointness assumption and no independence between membership indicators.

For an internal average $N(U) = N(U, U)$, each unordered edge appears twice in the ordered set $D(U, U)$. Both numerator and denominator contain this factor two, so the statistic equals the usual average over unordered internal edges. In concentration arguments the multiplicity is handled through $w_e = 2$ and $\sum_e w_e^2 = 2|D(U, U)|$. Hence the effective number of independent observations remains of order $|U|^2$, as required by the endpoint-search analysis.

Corollary B.13 (Scale form used below)**.** *Let $m \geq cn_\star\phi(\rho)$ and take $\lambda = c'\rho$. If*

$$sn_\star\phi(\rho)\rho^2 \geq C \log^2 n_\star,$$

then $\mathcal{E}_{\mathrm{avg}}(W, \lambda, m)$ holds with probability $1 - n_\star^{-\omega(1)}$.

Proof. Apply Lemma B.9 with $\varphi = m/n_\star$. The displayed condition dominates both the entropy term $\log(e/\varphi)$ and the extra $\log n_\star$ needed to turn the exponential failure probability into $n_\star^{-\omega(1)}$. Monotonicity of ϕ absorbs fixed rescalings of ρ into the constants. □

Why sample splitting is sufficient. Let $\mathcal{F}_j$ be the sigma-field generated by the latent sample and all edge variables revealed through stage j. If a candidate family $\mathcal{C}_j$ is $\mathcal{F}_j$-measurable and the next scores use only edges outside $\mathcal{F}_j$, then, conditionally on $\mathcal{F}_j$, every candidate is fixed and the new edges retain their product law. Lemmas B.5–B.9 therefore apply conditionally to $\mathcal{C}_j$. Taking expectations removes the conditioning. This argument permits adaptive selection, provided evaluation never reuses the edges that selected the candidate.

This distinction is essential. Uniform concentration over all m-subsets is affordable only on the small discovery block and is supplied by Lemma B.9. After the maximal separated family has been formed, navigation concentration is unioned only over the surviving seeds. The proof never assumes that an adaptively selected set is independent of the entire observed graph.

C INTERNAL-AVERAGE CANDIDATE CONSTRUCTION

C.1 ESTIMATING THE ENDPOINT AND OBTAINING A WITNESS

Fix $W \subseteq V$, $n_\star = |W|$, and a search scale $0 < \rho \le \min\{r_\mu, r_p/2\}$. Set

$$m_\rho := \left\lfloor \frac{\phi(\rho/3)}{2} n_\star \right\rfloor, \qquad \widehat{p}(0) := \max_{U \subseteq W,\ |U| = m_\rho} N(U).$$

Lemma C.1 (Endpoint estimation). *On $\mathcal{E}_{\mathrm{occ}}(W,\rho) \cap \mathcal{E}_{\mathrm{avg}}(W, L_p\rho, m_\rho)$, provided $m_\rho \ge 2$,*

$$p(0) - 3L_p\rho \le \widehat{p}(0) \le p(0) + L_p\rho.$$

Moreover, every $x \in \mathcal{M}$ has an m_ρ-subset $U_x \subseteq W \cap B(x,\rho)$ satisfying

$$N(U_x) \ge \widehat{p}(0) - 4L_p\rho.$$

Proof. Occupancy supplies $U_x \subseteq B(x,\rho)$. Every off-diagonal latent distance in U_x is below 2ρ, so

$$N(U_x) \ge p(U_x) - L_p\rho \ge p(2\rho) - L_p\rho \ge p(0) - 3L_p\rho.$$

This proves the lower bound and the covering assertion. Conversely, monotonicity gives $p(U) \le p(0)$ for every U, and uniform concentration gives $N(U) \le p(0) + L_p\rho$. Taking the maximum proves the upper bound. □

Define the observable candidate family

$$\mathcal{C}_\rho := \{U \subseteq W : |U| = m_\rho, \quad N(U) \ge \widehat{p}(0) - 4L_p\rho\}. \tag{27}$$

On the event of Lemma C.1, every $U \in \mathcal{C}_\rho$ satisfies

$$p(U) \ge p(0) - 8L_p\rho. \tag{28}$$

Lemma C.2 (Candidate coverage and observable slack). *On the event of Lemma C.1, the family $\mathcal{C}_\rho$ has the following properties.*

1. *For every $x \in \mathcal{M}$, it contains a set $U_x \subseteq W \cap B(x,\rho)$.*
2. *Every $U \in \mathcal{C}_\rho$ obeys* (28).
3. *If the empirical tolerance $4L_p\rho$ in* (27) *is replaced by any $a\rho$ with $a \ge 4L_p$, the same two conclusions hold with $8L_p$ replaced by $a + 4L_p$.*

Thus the numerical threshold is not tied to a unique constant; it only needs enough slack to cover one endpoint-estimation error and one pair-average error.

Proof. Lemma C.1 constructs U_x and proves $N(U_x) \ge \widehat{p}(0) - 4L_p\rho$, so $U_x \in \mathcal{C}_\rho$. For an arbitrary retained U, pair concentration gives

$$p(U) \ge N(U) - L_p\rho \ge \widehat{p}(0) - 5L_p\rho \ge p(0) - 8L_p\rho.$$

The last inequality uses the lower endpoint bound $\widehat{p}(0) \ge p(0) - 3L_p\rho$. With threshold $a\rho$, the identical calculation yields $p(U) \ge p(0) - (a + 4L_p)\rho$. The genuine ball candidates remain retained because $a \ge 4L_p$. □

The lemma separates two logically different uses of the internal optimization. Its maximum estimates the inaccessible value $p(0)$, while its near-maximizers create a covering family. The optimization need not return the unique best subset, and ties need not be resolved geometrically. Every near-maximizer receives its own witness in the next lemma, after which the net step removes redundancy.

Lemma C.3 (Radius witness from internal deficit). *Let $|U| \ge 2$ and suppose $p(U) \ge p(0) - \Delta$. There is a vertex $u_U \in U$ and a witness $\theta_U := X_{u_U}$ such that*

$$\frac{1}{|U|} \sum_{u \in U} \min\{d(X_u, \theta_U), r_p\} \le \frac{\Delta}{\ell_p}. \tag{29}$$

Consequently, for every $0 < R \le r_p$,

$$\frac{|\{u \in U : X_u \notin B(\theta_U, R)\}|}{|U|} \le \frac{\Delta}{\ell_p R}. \tag{30}$$

If U *is link-admissible, then* $\mathrm{rad}_{\mathrm{av}}(U, \theta_U) \le \Delta/\ell_p$.

Proof. Writing the ordered off-diagonal average by its first endpoint, at least one $u_U \in U$ obeys

$$\frac{1}{|U|-1} \sum_{u \in U \setminus \{u_U\}} \big[p(0) - p(d(X_u, X_{u_U}))\big] \le \Delta.$$

Monotonicity and the inverse local slope imply

$$p(0) - p(d(X_u, \theta_U)) \ge p(0) - p(\min\{d(X_u, \theta_U), r_p\}) \ge \ell_p \min\{d(X_u, \theta_U), r_p\}.$$

Adding the zero term for u_U gives (29). Markov's inequality gives (30). If U is link-admissible, truncation is inactive and the same estimate is the ordinary average-radius bound. □

Thus each candidate in (27) is an observable set with an analysis-only witness θ_U. The algorithm never evaluates θ_U or any latent distance.

C.2 Quantile consequences of the witness bound

The sliding estimate (30) is stronger than choosing a single radius. It provides a complete family of radius–contamination tradeoffs. The following formulation is used implicitly when the proof moves between rough candidates and approximate clusters.

Lemma C.4 (From internal deficit to approximate clusters)**.** *Under the assumptions of Lemma C.3, fix* $\eta \in (0, 1]$ *and put*

$$R_\eta := \min\left\{ r_p, \frac{\Delta}{\ell_p \eta} \right\}.$$

If $\Delta \le \ell_p \eta r_p$*, then* U *is an* (R_η, η)*-cluster centered at* θ_U*. In particular, if* $\Delta \le \ell_p \tau^2$*, then* U *is a* (τ, τ)*-cluster for every* $\tau \le r_p$*.*

Proof. The first numerical condition makes $R_\eta = \Delta/(\ell_p \eta) \le r_p$. Apply (30) at radius R_η to obtain

$$\frac{|\{u \in U : d(X_u, \theta_U) > R_\eta\}|}{|U|} \le \frac{\Delta}{\ell_p R_\eta} = \eta.$$

This is exactly the definition of an (R_η, η)-cluster. Setting $\eta = \tau$ and using $R_\tau \le \tau$ gives the final assertion. □

Lemma C.5 (Truncated layer-cake bound)**.** *Let* $R_U(u) = \min\{d(X_u, \theta_U), r_p\}$*. Then*

$$\frac{1}{|U|} \sum_{u \in U} R_U(u) = \int_0^{r_p} \tau_U(\theta_U; t)\, dt. \tag{31}$$

Consequently, the internal deficit controls not only one quantile but the integrated tail profile:

$$\int_0^{r_p} \tau_U(\theta_U; t)\, dt \le \frac{\Delta}{\ell_p}.$$

Proof. For every nonnegative $z \le r_p$, the identity $z = \int_0^{r_p} \mathbf{1}\{z > t\}\, dt$ holds. Apply it to $R_U(u)$, average over u, and interchange the finite sum and integral. The resulting integrand is precisely the fraction of vertices farther than t from θ_U. Lemma C.3 supplies the final bound. □

The symmetric choice $\eta = R = \sqrt{\Delta/\ell_p}$ explains the square-root scale without invoking a loose qualitative Markov argument: among all pairs (R, η) constrained by $R\eta \ge \Delta/\ell_p$, the maximum of R and η is minimized at equal values. Thus the unrestricted route naturally converts an $O(\rho)$ internal deficit into an $O(\sqrt{\rho})$ geometric radius and contamination level. The later link-safe route avoids this conversion by controlling the full average radius rather than balancing a radius against an uncontrolled tail.

D GEOMETRY OF ROUGH NEIGHBORHOOD CANDIDATES

D.1 SCORE APPROXIMATION AND THE SQUARE-ROOT LOSS

Lemma D.1 (Score transfer). *Let U satisfy the assumptions of Lemma C.3.*

1. If $U \cup \{y, \theta_U\}$ is link-admissible, then

$$|S_U(y) - p(d(\theta_U, y))| \le L_p \operatorname{rad}_{\mathrm{av}}(U, \theta_U) \le \frac{L_p}{\ell_p}\Delta.$$

2. Without link-admissibility, for every $0 < R \le r_p$ for which the near comparisons remain in $[0, r_p]$,

$$|S_U(y) - p(d(\theta_U, y))| \le L_p R + \frac{2M_p\Delta}{\ell_p R}. \tag{32}$$

In particular, choosing $R \asymp \sqrt{\Delta}$ gives score error $O(\sqrt{\Delta})$.

Proof. For the first claim, the local Lipschitz property and the triangle inequality give

$$|p(d(X_u, y)) - p(d(\theta_U, y))| \le L_p d(X_u, \theta_U).$$

Averaging and applying Lemma C.3 proves the claim.

For the second, divide U into points within R of θ_U and the remaining points. The former contribute at most $L_p R$ after averaging. The absolute contribution of the latter is at most $2M_p$ times their fraction, which is bounded by (30). This yields (32). Balancing the two terms proves the last statement. □

Lemma D.2 (Optimized robust score transfer). *Under the second setting of Lemma D.1, define*

$$A := \frac{2M_p\Delta}{\ell_p}.$$

If $\sqrt{A/L_p} \le r_p$ and the corresponding near comparisons stay in the link window, then

$$|S_U(y) - p(d(\theta_U, y))| \le 2\sqrt{AL_p} = 2\sqrt{\frac{2M_pL_p}{\ell_p}\Delta}. \tag{33}$$

If the unconstrained optimizer exceeds r_p, the valid bound is instead

$$L_p r_p + \frac{A}{r_p}.$$

Proof. The right-hand side of (32) is the convex function $f(R) = L_pR + A/R$ on $(0, r_p]$. Its unconstrained minimizer is $R_\star = \sqrt{A/L_p}$, where the two contributions are equal and $f(R_\star) = 2\sqrt{AL_p}$. If $R_\star > r_p$, convexity makes $R = r_p$ the constrained minimizer. Substitution proves both claims. □

This calculation isolates the source of the square root. It is not a concentration loss: even if every empirical score equaled its population value, the candidate may allocate a fraction $O(\Delta/R)$ of its mass beyond radius R. The local part contributes $O(R)$ Lipschitz bias and the uncontrolled part contributes $O(\Delta/R)$. With no certificate excluding the latter points from flat regions of the link, balancing the two terms is unavoidable for this argument. Once link safety is available, every point remains in the informative window, the ordinary average-radius bound applies, and the bias becomes $O(\Delta)$.

Lemma D.3 (Stability under a displaced witness). *Suppose $x, x' \in \mathcal{M}$ satisfy $d(x, x') \le h$, and all distances used below lie in $[0, r_p]$. Then for every nonempty U and every query y,*

$$|S_U(y) - p(d(x', y))| \le |S_U(y) - p(d(x, y))| + L_p h.$$

Moreover,

$$\operatorname{rad}_{\mathrm{av}}(U, x') \le \operatorname{rad}_{\mathrm{av}}(U, x) + h.$$

Proof. The first assertion follows by inserting $p(d(x,y))$ and using $|d(x,y) - d(x',y)| \le d(x,x')$. The second follows from $d(X_u, x') \le d(X_u, x) + h$ and averaging. □

This elementary stability fact is what allows a net of witnesses to stand in for arbitrary latent centers. The construction does not need a retained witness exactly at x; a witness within $O(r)$ changes every relevant score threshold by only $O(r)$.

D.2 Cross averages and link-admissible comparison

Lemma D.4 (Cross-average approximation inside the link window)**.** *Let U_1, U_2 be nonempty sets and let $x_1, x_2 \in \mathcal{M}$. Suppose*

$$\mathrm{rad}_{\mathrm{av}}(U_i, x_i) \le R, \qquad i = 1, 2,$$

and every distance among $X_{U_1} \cup X_{U_2} \cup \{x_1, x_2\}$ is at most r_p. If $|U_1|, |U_2| \ge 2$, then

$$|p(U_1, U_2) - p(d(x_1, x_2))| \le 4L_p R. \tag{34}$$

Proof. For every $(u_1, u_2) \in D(U_1, U_2)$, the two distances lie in the link window and

$$|d(X_{u_1}, X_{u_2}) - d(x_1, x_2)| \le d(X_{u_1}, x_1) + d(X_{u_2}, x_2).$$

After applying the upper Lipschitz inequality and averaging, each endpoint contribution is repeated at most the size of the opposite set. Since

$$|D(U_1, U_2)| = |U_1||U_2| - |U_1 \cap U_2| \ge \tfrac{1}{2}|U_1||U_2|,$$

the normalization costs at most a factor two. The two endpoint terms together yield (34). □

Lemma D.5 (Two-sided cross-score separation)**.** *Let U_i and x_i satisfy the assumptions of Lemma D.4 with average radii R_i, and suppose pair concentration has error at most λ. Put $d_{12} = d(x_1, x_2)$. Then*

$$p(d_{12}) - 2L_p(R_1 + R_2) - \lambda \le N(U_1, U_2) \le p(d_{12}) + 2L_p(R_1 + R_2) + \lambda. \tag{35}$$

Consequently, if $d_{12} \le a$ and $d'_{12} \ge b > a$ for two admissible candidate pairs with the same radius and concentration budgets, their observable cross scores are separated whenever

$$\ell_p(b - a) > 4L_p(R_1 + R_2) + 2\lambda. \tag{36}$$

Proof. Repeating the proof of Lemma D.4 without replacing both average radii by their common upper bound gives

$$|p(U_1, U_2) - p(d_{12})| \le 2L_p(R_1 + R_2).$$

The empirical event adds λ, proving (35). For the final assertion, monotonicity and the inverse local slope give $p(a) - p(b) \ge \ell_p(b - a)$. Subtract the two worst-case error budgets; condition (36) makes the remainder positive. □

Equation (36) is the quantitative core of every candidate adjacency rule below. The near and far geometric hypotheses are not tested directly. Instead, one chooses an observable threshold between their worst-case score ranges. The fixed constants K_{sep} and K_{cov} are simply large enough to make this interval nonempty after endpoint, geometric, and empirical errors have all been charged.

Lemma D.6 (Pair averages for contaminated local clusters)**.** *Let U_i be a (τ, τ)-cluster with center x_i, $i = 1, 2$, where $0 < \tau \le 1/2$ and $|U_1| = |U_2| = m \ge 2$. Put $d_{12} = d(x_1, x_2)$. Then*

$$p(d_{12} + 2\tau) - 8M_p\tau \le p(U_1, U_2) \le p((d_{12} - 2\tau)_+) + 8M_p\tau. \tag{37}$$

Proof. Choose good subsets $G_i \subseteq U_i$ with $|G_i| \ge (1 - \tau)m$ and $X_{G_i} \subseteq B(x_i, \tau)$. On $D(G_1, G_2)$, the triangle inequality gives

$$(d_{12} - 2\tau)_+ \le d(X_{u_1}, X_{u_2}) \le d_{12} + 2\tau.$$

The ordered pairs with at least one endpoint outside $G_1 \times G_2$ number at most $2\tau m^2$, whereas $|D(U_1, U_2)| \ge m(m-1) \ge m^2/2$. Their fraction is therefore at most 4τ. Replacing the corresponding link values, each bounded in absolute value by M_p, changes the average by at most $8M_p\tau$. Monotonicity of p on the remaining pairs proves both sides of (37). □

Corollary D.7 (Separated score levels for a link-safe candidate)**.** *Suppose $p(U) \geq p(0) - \Delta$ and U is λ-link-safe for some $\lambda < 1/4$. Let θ_U be the witness from Lemma C.3. For radii $0 < r < R$ satisfying*

$$r \leq (1-\lambda)r_p, \qquad R \leq (1-2\lambda)r_p,$$

we have

$$d(y, \theta_U) \leq r \Longrightarrow S_U(y) \geq p(0) - L_p r - \frac{L_p}{\ell_p}\Delta,$$

$$d(y, \theta_U) \geq R \Longrightarrow S_U(y) \leq p(0) - \ell_p R + \frac{L_p}{\ell_p}\Delta.$$

Hence a constant choice $R = Kr$ gives a score gap of order r whenever $\Delta = O(r)$ and K is sufficiently large.

Proof. Link safety and $\theta_U \in X_U$ imply $d(X_u, \theta_U) \leq \lambda r_p$ for every $u \in U$. Lemma C.3 gives average radius at most Δ/ℓ_p. If $d(y, \theta_U) \leq r$, then all $d(X_u, y)$ remain within r_p, so Lemma D.1 and the upper local slope give the first inequality.

For the far bound, first suppose $d(y, \theta_U) \leq (1-\lambda)r_p$; the same score approximation and the lower local slope apply. If $d(y, \theta_U) > (1-\lambda)r_p$, then every $u \in U$ obeys $d(X_u, y) > (1-2\lambda)r_p \geq R$. Monotonicity gives $S_U(y) \leq p(R) \leq p(0) - \ell_p R$, which is stronger. Subtracting the displayed levels proves the final assertion. □

Lemma D.8 (Score gap for an approximate neighborhood)**.** *Let U be a (τ, τ)-cluster with center x, where $0 < \tau \leq 1/2$ and $2\tau \leq r_p$. Then*

$$d(y, x) \leq \tau \quad \Longrightarrow \quad S_U(y) \geq p(0) - 2(L_p + M_p)\tau.$$

If $(K-1)\tau \leq r_p$ and $K \geq 1 + 8(L_p + M_p)/\ell_p$, then

$$d(y, x) \geq K\tau \quad \Longrightarrow \quad S_U(y) \leq p(0) - 4(L_p + M_p)\tau.$$

Proof. Choose $G \subseteq U$ with $|G| \geq (1-\tau)|U|$ and $X_G \subseteq B(x, \tau)$. If $d(y, x) \leq \tau$, then $d(X_u, y) < 2\tau$ for $u \in G$, and the lower Lipschitz estimate gives $p(d(X_u, y)) \geq p(0) - 2L_p\tau$. The remaining fraction is at most τ and contributes at worst $-M_p$, yielding the first bound.

If $d(y, x) \geq K\tau$, then $d(X_u, y) \geq (K-1)\tau$ on G. Monotonicity and the inverse local slope give

$$S_U(y) \leq p(0) - (1-\tau)\ell_p(K-1)\tau.$$

The stated choice of K and $\tau \leq 1/2$ imply the second bound. □

Lemma C.3 converts a deficit $\Delta = O(\rho)$ into a (τ, τ)-cluster at $\tau \asymp \sqrt{\rho}$. This is precisely the square-root loss in the unrestricted local-link route. If a candidate is link-safe, Corollary D.7 instead gives a score error linear in ρ. The distinction between these two implications is the reason for the coarse-to-fine bootstrap in Appendix H.

E Candidate nets from observable families

E.1 Candidate nets

The family $\mathcal{C}_\rho$ can be exponentially large. We reduce it before any full-sample operation. Join two candidates when their empirical cross-average is within a constant multiple of the search scale of $\widehat{p}(0)$, and retain a maximal independent set $\widehat{\mathcal{N}}_\rho$. In the link-safe route we first discard candidates not certified link-safe.

Lemma E.1 (Abstract observable-net principle)**.** *Let $\mathcal{C}$ be a finite family, let $\theta : \mathcal{C} \to \mathcal{M}$ assign an analysis-only witness, and let G be a graph on $\mathcal{C}$. Suppose there are numbers $0 < a < b$ such that*

$$d(\theta_U, \theta_{U'}) \leq a \Longrightarrow \{U, U'\} \in E(G), \tag{38}$$

$$\{U, U'\} \in E(G) \Longrightarrow d(\theta_U, \theta_{U'}) \leq b. \tag{39}$$

Assume also that for every $x \in \mathcal{M}$ there is $U_x \in \mathcal{C}$ with $d(x, \theta_{U_x}) \leq h$. If $\mathcal{I}$ is any maximal independent set of G, then

1. *distinct witnesses in $\theta(\mathcal{I})$ are more than a apart;*
2. *$\theta(\mathcal{I})$ is an $(h+b)$-net of $\mathcal{M}$;*
3. *if each witness is a distinct sampled vertex, then $|\mathcal{I}| \le |W|$.*

Proof. If two retained candidates had witness distance at most a, implication (38) would make them adjacent, contradicting independence. For any x, maximality implies that U_x is retained or adjacent to a retained U. In the first case $d(x, \theta_U) \le h$. In the second, (39) and the triangle inequality give

$$d(x, \theta_U) \le d(x, \theta_{U_x}) + d(\theta_{U_x}, \theta_U) \le h + b.$$

This proves covering. The last statement follows from the injection of retained candidates into their distinct witness vertices. □

The lemma clarifies why the adjacency rule is intentionally asymmetric in its interpretation. A near pair must be adjacent, because otherwise an independent set could retain two geometrically redundant seeds. Conversely, every observed adjacency must have a bounded geometric diameter, because maximality covers a discarded candidate only through one of its retained neighbors. Nothing is required of pairs at intermediate distances, and no transitivity of adjacency is used.

Lemma E.2 (Threshold stability for candidate graphs)**.** *Suppose a population cross score $Q(U, U')$ has a near lower bound q_{near} and a far upper bound q_{far}, with gap $g = q_{\mathrm{near}} - q_{\mathrm{far}} > 0$. If an observable score $\widehat{Q}$ satisfies $|\widehat{Q} - Q| \le \varepsilon$ uniformly and $2\varepsilon < g$, then every threshold*

$$T \in (q_{\mathrm{far}} + \varepsilon, q_{\mathrm{near}} - \varepsilon)$$

makes all near pairs adjacent under the rule $\widehat{Q} \ge T$ and all far pairs nonadjacent.

Proof. For a near pair, $\widehat{Q} \ge q_{\mathrm{near}} - \varepsilon > T$. For a far pair, $\widehat{Q} \le q_{\mathrm{far}} + \varepsilon < T$. The interval is nonempty because $2\varepsilon < g$. □

In the actual candidate graph, $Q = p(U_1, U_2)$, the empirical perturbation is supplied by $\mathcal{E}_{\mathrm{avg}}$, and the near–far population gap comes from Lemma D.5 or Lemma D.6. The endpoint estimate $\widehat{p}(0)$ perturbs the displayed threshold by another $O(\rho)$ term. Choosing the separation multiplier after all three budgets have been specified realizes the interval in Lemma E.2.

Proposition E.3 (Observable candidate net)**.** *There are constants $c, C, K > 0$ with the following property. On the events in Lemma C.1:*

1. *with a fuzzy link-safe certificate and $\rho = r$, the retained witnesses are cr-separated and form a Cr-net of $\mathcal{M}$;*
2. *without such a certificate, taking $\rho = cr^2$ makes every candidate an (r, r)-cluster, and the same cross-average rule produces witnesses that are cr-separated and form a Cr-net;*
3. *in both cases $|\mathcal{N}_\rho| \le |W|$.*

All membership and adjacency decisions depend only on observed edge averages.

Proof. We give the common argument and indicate where the two routes differ. For candidates U_i with witnesses θ_i, if the candidates are link-safe, Lemma D.1 gives

$$|p(U_1, U_2) - p(d(\theta_1, \theta_2))| \le C\rho.$$

Uniform pair concentration adds another $O(\rho)$ term. Therefore sufficiently close witnesses make the candidates adjacent, while the lower slope ℓ_p ensures that witnesses separated by a sufficiently large multiple of the threshold are nonadjacent. Constants are chosen so that all compared distances stay below r_p.

Without a certificate, (28) and Lemma C.3 imply

$$U \text{ is an } (r, r)\text{-cluster whenever } r^2 \ge C\rho.$$

For two such clusters, all but an $O(r)$ fraction of pairs have endpoints within r of their witnesses. The good pairs are controlled by the local Lipschitz bounds, and the bad pairs contribute at most $O(M_p r)$. Hence close witnesses are again adjacent and sufficiently distant witnesses are not. This proves separation of an independent set.

For covering, Lemma C.1 supplies, for every $x \in \mathcal{M}$, a genuine ball candidate $U_x \subseteq B(x, \rho)$. In the certified route it passes the link-safe filter when the scale is small enough. By maximality, U_x is retained or is adjacent to a retained candidate, and the preceding distance implication places the retained witness within Cr of x. Finally, separated witnesses are distinct sampled points, so $U \mapsto u_U$ injects $\mathcal{N}_\rho$ into W. □

E.2 The link-safe route in detail

Let $\mathcal{C}_\rho^{\mathrm{safe}}$ contain those $U \in \mathcal{C}_\rho$ whose internal pairs are all accepted by the fuzzy certificate. Choose a separation radius $\rho_{\mathrm{sep}} = K_{\mathrm{sep}}\rho$ and join $U_1, U_2 \in \mathcal{C}_\rho^{\mathrm{safe}}$ when

$$N(U_1, U_2) \geq \widehat{p}(0) - 2L_p\rho_{\mathrm{sep}}. \tag{40}$$

Lemma E.4 (Separation and covering in the link-safe route)**.** *For K_{sep} sufficiently large and ρ_{sep} sufficiently below r_p, the graph in* (40) *has the following deterministic properties on the good events of Lemma C.1:*

1. *$d(\theta_{U_1}, \theta_{U_2}) \leq \rho_{\mathrm{sep}}$ implies that U_1, U_2 are adjacent;*
2. *adjacency implies $d(\theta_{U_1}, \theta_{U_2}) \leq K_{\mathrm{cov}}\rho_{\mathrm{sep}}$, where K_{cov} depends only on L_p/ℓ_p;*
3. *a maximal independent set has separated witnesses and covers $\mathcal{M}$ at radius $(1 + K_{\mathrm{cov}})\rho_{\mathrm{sep}}$.*

Proof. For $U \in \mathcal{C}_\rho^{\mathrm{safe}}$, (28) and Lemma C.3 give

$$\mathrm{rad}_{\mathrm{av}}(U, \theta_U) \leq \frac{8L_p}{\ell_p}\rho.$$

The certificate gives a fixed margin to the boundary of the link window. Hence, whenever $d_{12} := d(\theta_{U_1}, \theta_{U_2}) \leq \rho_{\mathrm{sep}}$, Lemma D.4, pair concentration, and Lemma C.1 yield

$$\begin{aligned} N(U_1, U_2) &\geq p(d_{12}) - C\rho \\ &\geq p(0) - L_p\rho_{\mathrm{sep}} - C\rho \geq \widehat{p}(0) - 2L_p\rho_{\mathrm{sep}}, \end{aligned}$$

where the last inequality follows from $\rho_{\mathrm{sep}} = K_{\mathrm{sep}}\rho$ and a sufficiently large K_{sep}. This proves the first claim.

Conversely, suppose $d_{12} > K_{\mathrm{cov}}\rho_{\mathrm{sep}}$ but remains within the common link window. The same estimates, now using the inverse local slope, give

$$N(U_1, U_2) \leq p(0) - \ell_p d_{12} + C\rho < \widehat{p}(0) - 2L_p\rho_{\mathrm{sep}}$$

once $K_{\mathrm{cov}} > 1 + 2L_p/\ell_p$ and K_{sep} is large. If d_{12} is beyond the common link window, link safety implies that every cross-distance is at least a fixed positive fraction of r_p, so monotonicity and (14) produce an even larger drop. Thus adjacency forces the asserted distance bound.

Lemma C.1 supplies for every $x \in \mathcal{M}$ a candidate $U_x \subseteq B(x, \rho)$. It is certificate-safe when ρ is below the inner oracle margin, and its witness lies within ρ of x. Maximality of the independent set either retains U_x or makes it adjacent to a retained candidate. The second claim and the triangle inequality then give covering. Nonadjacency plus the first claim gives separation. □

E.3 The unrestricted local-link route in detail

Without a certificate, set $\rho = c_\rho r^2$. Lemma C.3 gives

$$\frac{|\{u \in U : X_u \notin B(\theta_U, r)\}|}{|U|} \leq \frac{8L_p\rho}{\ell_p r} \leq r$$

after choosing $c_\rho \leq \ell_p/(8L_p)$. Thus every candidate is an (r, r)-cluster. Join two candidates when

$$N(U_1, U_2) \geq \widehat{p}(0) - 2L_p r_{\mathrm{sep}}, \qquad r_{\mathrm{sep}} = K_{\mathrm{sep}} r. \tag{41}$$

Lemma E.5 (Separation and covering without a certificate)**.** *For constants $K_{\mathrm{sep}}, K_{\mathrm{cov}}$ depending only on the fixed link parameters, a maximal independent set in the graph* (41) *has r_{sep}-separated witnesses and covers $\mathcal{M}$ at radius $K_{\mathrm{cov}} r_{\mathrm{sep}}$.*

Proof. Lemma D.6 and uniform pair concentration give, for $d_{12} = d(\theta_{U_1}, \theta_{U_2})$,

$$N(U_1, U_2) \geq p(d_{12} + 2r) - 8M_p r - L_p \rho,$$
$$N(U_1, U_2) \leq p((d_{12} - 2r)_+) + 8M_p r + L_p \rho.$$

If $d_{12} \leq r_{\mathrm{sep}}$, the first inequality and the upper local slope show adjacency once K_{sep} dominates the contamination constants. If $d_{12} > K_{\mathrm{cov}} r_{\mathrm{sep}}$, the second inequality and the lower local slope put the cross-average below the threshold. The admissible range requirement $K_{\mathrm{cov}} r_{\mathrm{sep}} + 2r \leq r_p$ ensures that this inversion uses no information outside the local window.

The covering proof is identical to the final paragraph of Lemma E.4: every x has a genuine ball candidate U_x, and maximality links it to a retained candidate unless it is retained itself. The adjacency-to-distance implication then gives the desired covering radius. □

In both routes, separated witnesses are distinct sampled vertices. Consequently the number of retained candidates is at most $|W|$. This deterministic reduction is what makes later union bounds depend on the working-set size rather than on the number of enumerated subsets.

F Fresh-edge cluster refinement

F.1 A deterministic thresholding statement

Lemma F.1 (Thresholding with an arbitrary error set)**.** *Let $q : W \to \mathbb{R}$ be a population score, let $\widehat{q} : W \to \mathbb{R}$ be an observed score, and suppose there are sets*

$$A_{\mathrm{in}} \subseteq A_{\mathrm{out}} \subseteq W$$

and levels $q_{\mathrm{in}} > q_{\mathrm{out}}$ such that

$$q(v) \geq q_{\mathrm{in}} \quad (v \in A_{\mathrm{in}}), \qquad q(v) \leq q_{\mathrm{out}} \quad (v \notin A_{\mathrm{out}}).$$

Put $\Gamma = q_{\mathrm{in}} - q_{\mathrm{out}}$, choose $T = (q_{\mathrm{in}} + q_{\mathrm{out}})/2$, and define

$$I = \{v \in W : |\widehat{q}(v) - q(v)| > \Gamma/4\}, \qquad \widehat{A} = \{v : \widehat{q}(v) \geq T\}.$$

Then

$$A_{\mathrm{in}} \setminus I \subseteq \widehat{A} \subseteq A_{\mathrm{out}} \cup I. \tag{42}$$

If $|A_{\mathrm{in}}| \geq a$ and $|I| \leq b < a$, then

$$|\widehat{A}| \geq a - b, \qquad \frac{|\widehat{A} \setminus A_{\mathrm{out}}|}{|\widehat{A}|} \leq \frac{b}{a - b}.$$

Proof. If $v \in A_{\mathrm{in}} \setminus I$, then $\widehat{q}(v) \geq q_{\mathrm{in}} - \Gamma/4 > T$. If $v \notin A_{\mathrm{out}} \cup I$, then $\widehat{q}(v) \leq q_{\mathrm{out}} + \Gamma/4 < T$. These are the two inclusions in (42). The left inclusion gives the size bound. The right inclusion shows that every selected point outside A_{out} belongs to I, so division by the size lower bound proves the fraction estimate. □

This statement contains no probability and no geometry. Geometry supplies the nested sets $A_{\mathrm{in}} = W \cap B(x, r_{\mathrm{in}})$ and $A_{\mathrm{out}} = W \cap B(x, r_{\mathrm{out}})$; concentration bounds the error set I; occupancy lower-bounds $|A_{\mathrm{in}}|$. Separating these roles prevents the common logical mistake of treating threshold errors as if they were independent of the data-dependent seed.

F.2 One-pass and two-pass refinement

Lemma F.2 (Threshold refinement). *Let U be fixed before the edges from U to a fresh block W are revealed. Suppose that for some $x \in \mathcal{M}$, threshold level Θ, radii $r_{\mathrm{in}} < r_{\mathrm{out}}$, and $a_{\mathrm{in}} < a_{\mathrm{out}}$,*

$$d(y,x) \le r_{\mathrm{in}} \Longrightarrow S_U(y) \ge \Theta - a_{\mathrm{in}},$$
$$d(y,x) \ge r_{\mathrm{out}} \Longrightarrow S_U(y) \le \Theta - a_{\mathrm{out}}.$$

Let $\Gamma = a_{\mathrm{out}} - a_{\mathrm{in}} > 0$ and

$$T = \Theta - \frac{a_{\mathrm{in}} + a_{\mathrm{out}}}{2}, \qquad \widehat{U} = \{v \in W : N_U(v) \ge T\}.$$

Put

$$q = C\exp(-cs|U|\Gamma^2), \qquad b = 2q|W| + 2\log^2 N.$$

If $\mathcal{E}_{\mathrm{occ}}(W, r_{\mathrm{in}})$ holds and

$$b \le \frac{1}{4}|W|\phi(r_{\mathrm{in}}/3), \tag{43}$$

then, with conditional probability at least $1 - e^{-c\log^2 N}$, $|\widehat{U}| \ge |W|\phi(r_{\mathrm{in}}/3)/4$ and $\widehat{U}$ is an (r_{out}, η)-cluster with

$$\eta \le \frac{4b}{|W|\phi(r_{\mathrm{in}}/3)}. \tag{44}$$

If, more strongly, $|W|q \le N^{-\omega(1)}$, then with conditional probability $1 - N^{-\omega(1)}$,

$$W \cap B(x, r_{\mathrm{in}}) \subseteq \widehat{U} \subseteq W \cap B(x, r_{\mathrm{out}}), \tag{45}$$

and $|\widehat{U}| \ge |W|\phi(r_{\mathrm{in}}/3)/2$.

Proof. Let $I = \{v \in W : |N_U(v) - S_U(X_v)| > \Gamma/4\}$. For $v \notin I$ inside the inner ball,

$$N_U(v) \ge \Theta - a_{\mathrm{in}} - \Gamma/4 > T,$$

while for $v \notin I$ outside the outer ball,

$$N_U(v) \le \Theta - a_{\mathrm{out}} + \Gamma/4 < T.$$

The two pointwise implications can be summarized as

$$\big(W \cap B(x, r_{\mathrm{in}})\big) \setminus I \subseteq \widehat{U}, \qquad \widehat{U} \setminus B(x, r_{\mathrm{out}}) \subseteq I. \tag{46}$$

Corollary B.8 gives $|I| \le b$ except on an event of probability $e^{-c\log^2 N}$. On this event, occupancy and (43) imply

$$|\widehat{U}| \ge |W \cap B(x, r_{\mathrm{in}})| - |I| \ge \frac{1}{4}|W|\phi(r_{\mathrm{in}}/3).$$

The second inclusion in (46) then gives

$$\frac{|\widehat{U} \setminus B(x, r_{\mathrm{out}})|}{|\widehat{U}|} \le \frac{b}{|W|\phi(r_{\mathrm{in}}/3)/4},$$

which is (44).

Under the stronger assumption, a union bound and Lemma B.5 give $\mathbb{P}(I \ne \varnothing \mid X_U, X_W) \le |W|q = N^{-\omega(1)}$. Setting $I = \varnothing$ in (46) proves (45), and occupancy gives the sharper size lower bound. □

Proposition F.3 (Two-pass purification invariant). *Let $U^{(1)}$ be a seed fixed before edges to a fresh working block W_2 are observed. Suppose its population score has a gap $\Gamma_2 \ge cr$ between radii r and $K_2 r$. Let $U^{(2)}$ be the output of threshold refinement on W_2, and suppose the contamination conclusion gives*

$$|U^{(2)}| \ge c|W_2|\phi(cr), \qquad U^{(2)} \text{ is a } (K_2 r, c_2 r)\text{-cluster.}$$

If edges from $U^{(2)}$ to a second fresh block W_3 are independent of the first pass and

$$s|U^{(2)}|r^2 \ge C\log^2 n, \tag{47}$$

then a second threshold produces $U^{(3)}$ and a center x such that, with probability $1 - n^{-\omega(1)}$,

$$W_3 \cap B(x, K_2 r) \subseteq U^{(3)} \subseteq W_3 \cap B(x, K_3 r), \tag{48}$$

where K_3 depends only on the fixed model parameters.

Proof. Apply Lemma D.8 to $U^{(2)}$. Its radius and contamination are both $O(r)$, so after constants are absorbed into the notation the lemma gives a population gap $\Gamma_3 \geq c_3 r$ between an inner radius $K_2 r$ and an outer radius $K_3 r$. Conditional on the first pass, $U^{(2)}$ is fixed and the second-block edges remain independent. Lemma B.5 and (47) imply

$$\Pr\left\{ \max_{v \in W_3} |N_{U^{(2)}}(v) - S_{U^{(2)}}(X_v)| > \Gamma_3/4 \mid \mathcal{F}_2 \right\} \leq n^{1-cC} = n^{-\omega(1)}$$

after the logarithmic constant is enlarged. Thus the exceptional set in Lemma F.1 is empty. Its inclusions become exactly (48). □

The distinction between the two passes is statistical rather than geometric. The first pass tolerates an exceptional set and only needs enough accuracy to make its fraction $O(r)$. This output is already geometrically useful because Lemma D.8 is robust to an $O(r)$ fraction of contamination. The second pass spends the stronger $\log^2 n$ uniform budget on a much smaller seed family and eliminates every exception. A single pass could be made exact by paying the stronger budget immediately for every enumerated candidate, but that would destroy the intended separation between exhaustive search and full-sample work.

G THREE-BLOCK NEIGHBORHOOD EXTRACTION

G.1 OBSERVABLE PROCEDURE AND BLOCK ROLES

For clarity, we describe the extraction map before proving its guarantees. The procedure receives only the observed variables, fixed model constants, a target radius r, and optionally a fuzzy certificate. Latent witnesses and centers appear only in the proof.

1. **Discovery.** Choose a working subset $V_1' \subseteq V_1$. Enumerate all subsets of the prescribed size m_ρ, compute their internal averages, form $\widehat{p}(0)$ and $\mathcal{C}_\rho$, and build the observable candidate graph. Retain a maximal independent set $\mathcal{N}_\rho$.
2. **Contamination reduction.** Choose $V_2' \subseteq V_2$. For every $U \in \mathcal{N}_\rho$, compute the navigation scores $N_U(v)$ for $v \in V_2'$ using only cross-block edges. Threshold them at the midpoint of the deterministic score gap and obtain $U^{(2)}$.
3. **Exact refinement.** For every surviving $U^{(2)}$, compute fresh scores on V_3 and threshold a second time. This produces $U^{(3)}$ with exact outer-radius control.
4. **Assignment.** For each $v \in V_3$, select the first refined cluster whose certified inner ball contains v. The proof of net coverage guarantees existence. Add v if necessary; the radius constant absorbs this convention. Return the assigned set as U_v.

Every operation in this list is measurable with respect to observed edge variables. The phrase "whose inner ball contains v" in the last item describes the proof of existence; operationally the assignment uses the same passing score inequality that defines the refinement. The exact-refinement inclusion shows that a vertex passing that test is inside the claimed outer ball.

Definition G.1 (Stage sigma-fields)**.** Let $\mathcal{F}_1$ contain the latent sample and all edges internal to V_1'. Let $\mathcal{F}_2$ additionally contain the candidate-graph edges and the edges from retained seeds to V_2'. Let $\mathcal{F}_3$ additionally contain edges from the first-refinement outputs to V_3. The construction is arranged so that

$$\mathcal{N}_\rho \in \mathcal{F}_1, \qquad \{U^{(2)} : U \in \mathcal{N}_\rho\} \in \mathcal{F}_2, \qquad \{U^{(3)}\} \in \mathcal{F}_3,$$

and the edge variables first used at stage $j+1$ are independent of $\mathcal{F}_j$ conditional on the latent sample.

The construction can realize these sigma-fields either by disjoint vertex blocks, as stated here, or by an independent random coloring of unordered pairs into edge blocks. Vertex splitting is easier to state and loses only fixed factors in sample size. Edge coloring can reuse vertices more efficiently but does not change the proof.

G.2 BUDGET IMPLICATIONS

Lemma G.2 (Master scale implies all three stage budgets)**.** *Assume* (53)*, let n_1 and n_2 be the smallest admissible working-set sizes in* (54) *and* (55)*, and suppose $\rho \le r$. After fixed rescaling of arguments of ϕ, the following hold:*

$$n_1 \phi(c\rho) \ge C \log^2 n_1, \tag{49}$$

$$s n_1 \phi(c\rho) \rho^2 \ge C \log^2 n_1, \tag{50}$$

$$s n_2 \phi(cr) r^2 \ge C \log^2 n, \tag{51}$$

$$sn\phi(cr) r^2 \ge C \log^2 n. \tag{52}$$

The first two guarantee discovery occupancy and pair concentration; the third controls the exceptional fraction in the first refinement; the last makes the second refinement exact.

Proof. Equation (50) is the defining property of n_1. Since $s\rho^2 \le 1$, it implies (49). Feasibility of $n_1 \le n$ follows from (53). The definition of n_2 gives (51), and feasibility again follows from the master condition after monotonicity of ϕ and fixed rescaling. The final display is the target-scale part of the same master condition. Constants account for the fact that each vertex block has only a fixed fraction of the n vertices. □

Lemma G.3 (Event audit for one extraction run)**.** *Let $\mathcal{G}_1$ be the intersection of discovery occupancy and uniform pair concentration, let $\mathcal{G}_2$ be the first-refinement occupancy and exceptional-count event, and let $\mathcal{G}_3$ be the final occupancy and uniform navigation event. Under Lemma G.2,*

$$\Pr(\mathcal{G}_1^c) = n_1^{-\omega(1)}, \qquad \Pr(\mathcal{G}_2^c \mid \mathcal{F}_1) = n^{-\omega(1)}, \qquad \Pr(\mathcal{G}_3^c \mid \mathcal{F}_2) = n^{-\omega(1)}.$$

Consequently $\Pr(\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3) = 1 - o(1)$.

Proof. The first claim follows from Lemma B.2 and Corollary B.13 on V_1'. Conditional on $\mathcal{F}_1$, the retained family contains at most n_1 fixed seeds. Apply Corollary B.8 to each seed and union-bound; the $e^{-c\log^2 n}$ tail dominates the factor n_1. Conditional on $\mathcal{F}_2$, repeat the argument with the exact clause of Lemma F.2 on the full output block. Taking expectations of the conditional failure bounds and summing the three terms proves the last assertion. □

Proposition G.4 (Three-block extraction)**.** *Split the data into independent vertex blocks V_1, V_2, V_3 of comparable size. At target radius r, assume either:*

1. *a fuzzy link-safe certificate is available on V_1 and the search scale is $\rho = r$; or*
2. *no certificate is available and the search scale is $\rho = cr^2$.*

If

$$sn\phi(c\rho)\rho^2 \ge C \log^2 n \tag{53}$$

and the scales remain inside r_μ and r_p, then with probability $1 - o(1)$ the procedure returns, for every $v \in V_3$, a set $U_v \subseteq V_3$ such that

$$v \in U_v, \qquad X_{U_v} \subseteq B(X_v, Cr), \qquad |U_v| \ge cn\phi(cr).$$

Proof. **Stage 1: discovery on a small working set.** Choose $V_1' \subseteq V_1$ of the smallest size $n_1 \ge \log\log n$ for which

$$s n_1 \phi(c\rho)\rho^2 \ge C \log^2 n_1. \tag{54}$$

The assumed condition at size n guarantees feasibility. Lemmas B.2 and B.9 make

$$\mathcal{E}_{\mathrm{occ}}(V_1', \rho) \cap \mathcal{E}_{\mathrm{avg}}(V_1', L_p\rho, m_\rho)$$

hold with probability $1 - n_1^{-\omega(1)}$. On this event, Lemma C.1 constructs $\widehat{p}(0)$ and the candidate family $\mathcal{C}_\rho$ using only internal edges of V_1'. Apply Proposition E.3. In the certified route this invokes Lemma E.4; in the unrestricted route it invokes Lemma E.5. In either case the selected family $\mathcal{N}_\rho$ has at most n_1 members and its witnesses form a $C_0 r$-net of $\mathcal{M}$.

Stage 2: contamination reduction. Choose $V_2' \subseteq V_2$ of the smallest size n_2 satisfying

$$sn_2\phi(cr)r^2 \geq C\log^2 n. \tag{55}$$

Fix $U \in \mathcal{N}_\rho$ and write $x_U = \theta_U$. In the certified route, Corollary D.7 gives constants $a_{\mathrm{in}} < a_{\mathrm{out}}$ and radii r and $K_2 r$ such that

$$S_U(y) \geq \widehat{p}(0) - a_{\mathrm{in}} \quad \text{inside } B(x_U, r), \qquad S_U(y) \leq \widehat{p}(0) - a_{\mathrm{out}} \quad \text{outside } B(x_U, K_2 r),$$

with $\Gamma_2 = a_{\mathrm{out}} - a_{\mathrm{in}} \geq c_2 r$. In the unrestricted route the seed is an (r, r)-cluster, so Lemma D.8 gives the same conclusion after increasing K_2.

The seed size satisfies $|U| \asymp n_1\phi(c\rho)$. Combining (54) with $\Gamma_2 \gtrsim r$ makes the exceptional probability $q(|U|, \Gamma_2)$ small. Condition (55) makes the logarithmic part of the budget in (43) smaller than $c\phi(r/3)n_2$. Therefore Lemma F.2, in its contamination form, returns a set $U^{(2)} \subseteq V_2'$ satisfying

$$|U^{(2)}| \geq cn_2\phi(cr), \qquad U^{(2)} \text{ is a } (K_2 r, \eta_2)\text{-cluster}, \qquad \eta_2 \leq K_2 r.$$

Since $|\mathcal{N}_\rho| \leq n_1 \leq n$, the failure probability $e^{-c\log^2 n}$ can be unioned over all seeds.

Stage 3: exact refinement on the full output block. Absorb constants by calling every $U^{(2)}$ a (r_2, r_2)-cluster with $r_2 = K_2 r$. Lemma D.8, applied with inner radius r_2 and outer radius $r_3 = K_3 r_2$, supplies a new score gap $\Gamma_3 \geq c_3 r$. Moreover,

$$s|U^{(2)}|\Gamma_3^2 \gtrsim sn_2\phi(cr)r^2 \geq C\log^2 n.$$

Thus $nq(|U^{(2)}|, \Gamma_3) = n^{-\omega(1)}$, and the exact clause of Lemma F.2 applies to the fresh edges from V_2' to V_3. It gives

$$V_3 \cap B(x_U, r_2) \subseteq U^{(3)} \subseteq V_3 \cap B(x_U, r_3)$$

simultaneously over all retained seeds. The witness family is a $C_0 r$-net; choose constants so that $C_0 r < r_2$. Every X_v, $v \in V_3$, then lies in the inner ball of some refined seed and hence belongs to its output. Assign one such $U^{(3)}$ to v and call it U_v. Since $v \in U_v \subseteq B(x_U, r_3)$,

$$X_{U_v} \subseteq B(x_U, r_3) \subseteq B(X_v, 2r_3) \subseteq B(X_v, Cr).$$

Occupancy in $B(x_U, r_2)$ gives $|U_v| \geq cn\phi(cr)$. This proves the geometric and size conclusions.

Probability accounting. The discovery event fails with probability $n_1^{-\omega(1)} = o(1)$. The first-refinement occupancy and score events fail with probability $n_2^{-\omega(1)} + n^{-\omega(1)}$, and the final events fail with probability $n^{-\omega(1)}$. Each stage uses a fresh edge block; hence all conditional statements remain valid after conditioning on the preceding output. A union bound proves simultaneous success with probability $1 - o(1)$. □

The uncertified route in Proposition G.4 pays $\rho \asymp r^2$. The following bootstrap removes that loss while using only observations.

H Link-safe bootstrap and proof of Theorem 3.1

H.1 Geometry of coarse probes

The bootstrap does not estimate coordinates or the link globally. It uses a coarse exact cluster as a repeated probe for one local yes–no question: is a fresh vertex sufficiently close to the probe center? Overlapping affirmative regions then certify that two fresh vertices are mutually close.

Lemma H.1 (Probe-neighborhood sandwich)**.** *Let A be an exact cluster satisfying*

$$X_A \subseteq B(x, h), \qquad |A| \geq m, qquad|\widehat{p}_0 - p(0)| \leq e_0.$$

Fix radii $0 < a < b$ such that $h + b \leq r_p$, and let

$$\widehat{W}_A(T) = \{v : N_A(v) \geq \widehat{p}_0 - T\}.$$

Suppose the fresh navigation errors are at most e_{nav}. If

$$L_p(h + a) + e_0 + e_{\mathrm{nav}} < T < \ell_p(b - h) - e_0 - e_{\mathrm{nav}}, \tag{56}$$

then

$$B(x, a) \cap V_{\mathrm{fresh}} \subseteq \widehat{W}_A(T) \subseteq B(x, b) \cap V_{\mathrm{fresh}}. \tag{57}$$

Proof. If $d(X_v, x) \le a$, then every $u \in A$ satisfies $d(X_u, X_v) \le h + a \le r_p$. The upper local slope gives

$$S_A(X_v) \ge p(0) - L_p(h + a).$$

After subtracting the navigation error and replacing $p(0)$ by $\widehat{p}_0$, the left inequality in (56) gives $N_A(v) > \widehat{p}_0 - T$, so v is retained.

If $d(X_v, x) \ge b$, then every cluster point is at distance at least $b - h$. Monotonicity and the inverse local slope give

$$S_A(X_v) \le p(0) - \ell_p(b - h).$$

Adding both error terms and using the right inequality in (56) puts $N_A(v)$ below threshold. This proves the outer inclusion. □

The interval in (56) is nonempty whenever

$$\ell_p(b - h) - L_p(h + a) > 2(e_0 + e_{\mathrm{nav}}).$$

This is the precise coarse-resolution requirement. The cluster radius h, endpoint error e_0, and navigation error must all be small compared with the fixed link-window margin $b - a$. No target-scale accuracy is required at this stage.

Lemma H.2 (Shared probes give a fuzzy certificate)**.** *Suppose centers $\{x_z\}$ form an h-net and their observable probe sets satisfy*

$$B(x_z, a) \cap V_{\mathrm{fresh}} \subseteq W_z \subseteq B(x_z, b) \cap V_{\mathrm{fresh}}$$

with $h < a/4$. Declare a pair (v, w) accepted when $v, w \in W_z$ for some z. Then

$$d(X_v, X_w) \le a/2 \Longrightarrow \textit{accepted} \Longrightarrow d(X_v, X_w) \le 2b.$$

Proof. For the second implication, a shared probe and the outer inclusion give $d(X_v, X_w) \le d(X_v, x_z) + d(x_z, X_w) \le 2b$. For the first, choose a center with $d(X_v, x_z) \le h$. Then

$$d(X_w, x_z) \le d(X_w, X_v) + d(X_v, x_z) \le a/2 + h < a.$$

Both vertices belong to the inner ball and therefore to W_z. □

The certificate is deliberately fuzzy: pairs in the annulus between $a/2$ and $2b$ may be accepted or rejected. All later arguments use only the two displayed implications, so no attempt is made to classify that gray zone. This avoids demanding target-scale accuracy from a coarse round whose sole purpose is to keep subsequent comparisons inside the informative part of the link.

Lemma H.3 (Coarse clusters construct a fuzzy link-safe certificate)**.** *Suppose a coarse run at scale r_0 returns an estimate $\widehat{p}_0$ and exact clusters A_z centered at X_z, with*

$$|\widehat{p}_0 - p(0)| \le Cr_0, \qquad X_{A_z} \subseteq B(X_z, Cr_0), \qquad |A_z| \ge cn\phi(cr_0),$$

and the centers $\{X_z\}$ forming a Cr_0-net. If $r_0 \le cr_p$ and

$$sn\phi(cr_0)r_p^2 \ge C \log^2 n,$$

then fresh averages $N_{A_z}(v)$ construct, with probability $1 - n^{-\omega(1)}$, an observable fuzzy link-safe certificate on a fresh block.

Proof. Put

$$\alpha := \frac{\ell_p}{128 L_p}, \qquad \lambda := \frac{1}{8}, \qquad T_{\mathrm{safe}} := \widehat{p}_0 - \frac{\ell_p r_p}{32}.$$

For each coarse center z, use fresh edges to define the observable probe neighborhood

$$W_z := \{v : N_{A_z}(v) \ge T_{\mathrm{safe}}\}. \tag{58}$$

The assumed size of A_z and Lemma B.5 give, simultaneously over all z and all fresh vertices v,

$$|N_{A_z}(v) - S_{A_z}(X_v)| \le \varepsilon_{\mathrm{nav}} := C \frac{\log n}{\sqrt{sn\phi(cr_0)}}. \tag{59}$$

The scale condition makes $\varepsilon_{\mathrm{nav}} \le c\ell_p r_p$. We now establish two deterministic inclusions.

If $d(X_v, X_z) \le 2\alpha r_p$, then for every $u \in A_z$,

$$d(X_u, X_v) \le Cr_0 + 2\alpha r_p \le r_p.$$

The upper local slope therefore gives

$$S_{A_z}(X_v) \ge p(0) - L_p C r_0 - 2L_p \alpha r_p.$$

Combining this with $|\widehat{p}_0 - p(0)| \le Cr_0$ and (59), and taking r_0/r_p sufficiently small, yields $N_{A_z}(v) \ge T_{\mathrm{safe}}$. Thus

$$B(X_z, 2\alpha r_p) \cap V_{\mathrm{fresh}} \subseteq W_z. \tag{60}$$

Conversely, if $d(X_v, X_z) \ge r_p/16$, then

$$d(X_u, X_v) \ge r_p/16 - Cr_0 \qquad (u \in A_z).$$

Monotonicity and the inverse local slope imply

$$S_{A_z}(X_v) \le p(0) - \ell_p(r_p/16 - Cr_0).$$

After the same endpoint and navigation errors are added, the right-hand side remains strictly below T_{safe}. Hence

$$W_z \subseteq B(X_z, r_p/16). \tag{61}$$

Declare $\mathcal{O}_{\mathrm{safe}}(v, w) = 1$ if and only if $v, w \in W_z$ for at least one z. If it accepts, (61) and the triangle inequality give $d(X_v, X_w) < r_p/8 = \lambda r_p$. Conversely, if $d(X_v, X_w) \le \alpha r_p$, choose a net center with $d(X_v, X_z) \le Cr_0 \le \alpha r_p/4$. Then

$$d(X_w, X_z) \le d(X_w, X_v) + d(X_v, X_z) \le \frac{3}{2}\alpha r_p.$$

Both vertices therefore fall under (60). They share W_z and are accepted. The construction is a fuzzy certificate with fixed margins depending only on L_p/ℓ_p. □

Proof of Theorem 3.1. Use seven comparable vertex blocks. Set $r_0 = \sqrt{r}$ and apply Proposition G.4 without a certificate to the first three blocks at output scale r_0. Its internal discovery scale is $cr_0^2 = cr$, so its condition is

$$sn\phi(cr)r^2 \ge C \log^2 n,$$

which is exactly Equation (7) up to constants. The coarse output satisfies Lemma H.3; the same scale condition implies its navigation requirement because $r_0 \ge r$ and $r_0^2 = r$.

We spell out the two rounds because they perform logically different tasks.

Round 1: coarse geometric anchors. Let $\mathcal{A}_0 = \{A_z : z \in \mathcal{Z}_0\}$ be the family produced by the uncertified route. On its success event,

$$X_{A_z} \subseteq B(X_z, C_0 r_0), \qquad |A_z| \ge c_0 n\phi(c_0 r_0), \tag{62}$$

and $\{X_z : z \in \mathcal{Z}_0\}$ is a $C_0 r_0$-net of the latent points in the fourth block. The endpoint estimate made in the discovery step also satisfies $|\widehat{p}_0 - p(0)| \le C_0 r_0$. These conclusions are deliberately coarse: their purpose is not yet to approximate target-scale balls, but to supply repeated probes distributed throughout the sampled space.

Apply Lemma H.3 using only edges from the coarse anchors to the fourth block. Conditional on the first-round sigma-field, the sets A_z are fixed and those edges are independent. Hence the navigation bound (59) is valid without a union bound over all subsets that could have appeared during discovery. The result is a random function $\mathcal{O}_{\mathrm{safe}}$ satisfying, simultaneously for fourth-block vertices,

$$d(X_v, X_w) \le \alpha r_p \Longrightarrow \mathcal{O}_{\mathrm{safe}}(v, w) = 1 \Longrightarrow d(X_v, X_w) \le \lambda r_p. \tag{63}$$

The first implication prevents false negatives on genuinely local comparisons; the second prevents the fine-scale search from applying the inverse slope of p outside its stated range.

Round 2: target-scale extraction. Apply the certified route of Proposition G.4 to the last three blocks at scale r, using the certificate in (63). In this route the internal deficit is $O(r)$ rather than the

$O(r^2)$ deficit used by the uncertified route. The certificate makes every comparison link-admissible, so Corollary D.7 converts this deficit directly into an $\Omega(r)$ score margin. Consequently, the discovery, first-refinement, and exact-refinement conditions all reduce to

$$sn\phi(cr)r^2 \geq C \log^2 n. \tag{64}$$

For each target vertex v in the final block the proposition returns a seed $z(v)$ and a set U_v such that

$$v \in U_v, \qquad X_{U_v} \subseteq B(X_v, Cr), \qquad |U_v| \geq cn\phi(cr). \tag{65}$$

Replacing c inside $\phi(cr)$ by a smaller universal constant and using monotonicity of ϕ gives the theorem's stated form.

Uniform coverage and probability. The preceding construction treats one of the seven blocks as the final target block. Repeat it for a constant collection of cyclic assignments so that every original vertex appears once as a target. This does not mean rerunning an adaptive step on its own evaluation edges: within every assignment the seven roles remain edge-disjoint. Intersections of the uniform occupancy event, the two discovery events, the certificate event, and the four refinement events have probability

$$1 - O(n^{-A})$$

for every fixed A, after the constants in (7) are enlarged. There are only constantly many assignments, so a final union bound preserves $1 - n^{-\omega(1)}$. On that event (65) holds simultaneously for all vertices. Constants absorb the seven-way block-size loss.

Enumeration and runtime. The only exponential operation is the internal-average search on a small working block. It enumerates $m_1 \asymp n_1\phi(cr)$ subsets of a set of size n_1. The least admissible n_1 satisfies

$$sn_1\phi(cr)r^2 \gtrsim \log^2 n_1.$$

Write $q = \phi(cr)$ and take the smallest dyadic n_1 satisfying this inequality. The minimality of n_1, together with $\log n_1 \lesssim \log(e/q)/(sr^2)$ in the nontrivial regime, gives

$$n_1 q \log(e/q) \leq C\frac{\log^2(e/q)}{sr^2}.$$

The entropy estimate $\binom{n_1}{m_1} \leq \exp\{n_1 H(m_1/n_1)\}$ and $H(q) \leq q\log(e/q)$ therefore yield the conservative factor

$$\exp\left\{C\frac{\log^2(1/\phi(r))}{sr^2}\right\}.$$

The replacement of $\phi(cr)$ by $\phi(r)$ changes constants under the fixed scale convention used in the theorem; one may retain $\phi(cr)$ throughout when no regular comparison between the two values is assumed.

Let K_r denote this enumeration factor. The maximal independent-set computation uses at most K_r^2 scalar cross-comparisons on the working set, which is absorbed by enlarging the constant in the exponent. After separation, every seed is represented by one of at most K_r enumerated candidates. Navigation and refinement evaluate $O(\log^2 n)$ empirical averages per vertex and surviving seed, for total cost $O(K_r n \log^2 n)$. Endpoint estimation and cluster-to-cluster averaging are no larger. Since a constant number of block assignments only changes the leading constant, the total runtime and number of distinct output clusters obey the claimed bounds. □

I End-to-end invariant and parameter audit

This section collects the quantitative implications used across several earlier proofs. It adds no assumption to Theorem 3.1; its purpose is to make the dependency chain checkable from the master scale condition without requiring the reader to reconstruct constant choices from different sections.

I.1 THE INVARIANT CARRIED BY THE ALGORITHM

For a seed family $\mathcal{A}$, write $\Theta(\mathcal{A})$ for its analysis-only witness set. At the successive stages, the proof establishes the following invariant:

$$\begin{aligned}
\text{candidates:} &\quad \Delta_{\text{int}} = O(\rho), && \text{coverage by genuine ball candidates;}\\
\text{candidate net:} &\quad |\mathcal{N}_\rho| \le n_1, && O(r)\text{-separated and } O(r)\text{-covering witnesses;}\\
\text{first refinement:} &\quad U^{(2)} \text{ is a } (K_2 r, O(r))\text{-cluster}, && \text{with the same covering witnesses;}\\
\text{exact refinement:} &\quad X_{U^{(3)}} \subseteq B(\theta_U, K_3 r), && \text{with at most } n_1 \text{ distinct outputs;}\\
\text{assignment:} &\quad X_{U_v} \subseteq B(X_v, Cr), && \text{covering every target vertex.}
\end{aligned} \tag{66}$$

The first row is established by Lemmas C.2 and C.3. The second is Proposition E.3; the third and fourth are Lemma F.2 and Proposition F.3; the last uses the inner inclusion and witness-net coverage.

Lemma I.1 (Preservation of covering through refinement)**.** *Suppose $\Theta(\mathcal{N})$ is a hr-net and, for every $U \in \mathcal{N}$, an exact refinement $R(U)$ satisfies*

$$V_{\text{out}} \cap B(\theta_U, h'r) \subseteq R(U) \subseteq V_{\text{out}} \cap B(\theta_U, Hr),$$

where $h' > h$. Then every $v \in V_{\text{out}}$ belongs to at least one $R(U)$, and any such assigned cluster obeys

$$X_{R(U)} \subseteq B(X_v, 2Hr) \tag{67}$$

after increasing H so that $H \ge h'$.

Proof. Choose U with $d(X_v, \theta_U) \le hr < h'r$. The inner inclusion gives $v \in R(U)$. For any $w \in R(U)$, the outer inclusion and $v \in R(U)$ yield

$$d(X_w, X_v) \le d(X_w, \theta_U) + d(\theta_U, X_v) \le Hr + h'r \le 2Hr.$$

This proves both membership and (67). □

The lemma explains why refinement never needs to move or estimate a center. The same latent witness selected during internal-average discovery remains valid throughout the proof. Refinement changes membership, not the analytical center. In the final output the target vertex itself becomes the stated center by a single triangle-inequality step.

I.2 SCALE SUBSTITUTION IN THE TWO ROUNDS

Lemma I.2 (Coarse-to-fine scale matching)**.** *Let the desired final radius be $r < 1$ and set $r_0 = \sqrt{r}$. In the uncertified route with coarse output radius r_0, take internal search scale $\rho_0 = cr_0^2 = cr$. In the certified fine route, take $\rho_1 = cr$. Then both discovery conditions have the same order:*

$$sn\phi(c'r)r^2 \ge C \log^2 n. \tag{68}$$

The coarse probe condition $sn\phi(cr_0)r_p^2 \ge C \log^2 n$ also follows from (68) *after constants are adjusted and the target scale is restricted to the stated admissible range.*

Proof. For the coarse run, Proposition G.4 is applied with target radius r_0 and uncertified search scale $\rho_0 = cr_0^2$. Its concentration gap is of order ρ_0, so the discovery budget is

$$sn\phi(c\rho_0)\rho_0^2 \asymp sn\phi(c'r)r^2.$$

For the fine run the certificate removes the square-root conversion, and both target and search scales are of order r. The same expression results. Finally, $r_0 \ge r$ and monotonicity gives $\phi(cr_0) \ge \phi(c'r)$ for compatible constants. Since r_p is fixed and the admissible target range makes $r \le cr_p$, the master lower bound with factor r^2 is at least as strong as the fixed-window probe requirement after enlarging its constant. □

The identity $r_0^2 = r$ is therefore not a heuristic choice. If the coarse output scale were larger, its internal search would use a scale larger than r and would not resolve enough geometry to create a target-scale certificate. If it were smaller, the uncertified discovery budget would be more demanding than the fine round and would determine the overall rate. The square root equalizes the two costs.

I.3 A DETAILED ENTROPY AUDIT

Let $q = \phi(cr)$ and suppose the discovery working set has size n_1. Candidate enumeration considers subsets of size $m_1 = \lceil qn_1 \rceil$. Standard entropy bounds give

$$\log \binom{n_1}{m_1} \le n_1 H(m_1/n_1) \le C n_1 q \log(e/q). \tag{69}$$

The working-set condition requires

$$sn_1 qr^2 \ge C \log^2 n_1. \tag{70}$$

Choose the least dyadic n_1 satisfying (70). Unless n_1 is at its minimum value $\lceil \log \log n \rceil$, the preceding dyadic value fails, which bounds $n_1 qr^2 s$ above by a constant multiple of $\log^2 n_1$. Solving this relation coarsely gives

$$\log n_1 \le C \log(e/q)/(sr^2)$$

in the shrinking-scale regime. Substitution into (69) yields

$$\log \binom{n_1}{m_1} \le C \frac{\log^2(e/q)}{sr^2}. \tag{71}$$

This estimate is intentionally conservative. It upper-bounds the cost of listing all candidates, even though many fail the internal-average threshold immediately. The candidate graph can have quadratically many pair comparisons in the number enumerated, but squaring an exponential only doubles its exponent and is absorbed into C. After maximal-independent-set reduction, there are at most n_1 seeds; all later stages cost at most

$$Cnn_1 \operatorname{polylog} n$$

score operations for one block assignment. Bounding n_1 by the same exponential factor and using constantly many assignments gives the runtime in Theorem 3.1.

At fixed accuracy, q, s, and r are positive constants. Then the left-hand side of (70) is linear in n_1, while the right-hand side is only logarithmic squared, so the imposed floor $n_1 \asymp \log \log n$ is admissible for large n. Candidate enumeration is a fixed-base exponential in $\log \log n$, hence polynomial in $\log n$. The refinement working set has size $n_2 = O(\log^2 n)$, and the final block is processed only against the reduced seed family. This recovers the near-linear fixed-resolution conclusion without invoking any additional geometric structure.

I.4 FAILURE-PROBABILITY AUDIT

Proposition I.3 (Fixed-confidence form)**.** *Fix $A > 0$. After changing constants depending on A, the conclusion of Theorem 3.1 holds with probability at least $1 - n^{-A}$ under*

$$sn\phi(cr)r^2 \ge C_A \log n.$$

The runtime bound in Theorem 3.1 remains a valid conservative upper bound.

Proof. In Lemma B.6, take $u = (A + 12) \log n$ instead of $u = \log^2 n$. In Corollary B.8, the same Chernoff calculation replaces the additive $2 \log^2 N$ by $C_A \log N$ and the tail by N^{-A-12}. The uniform occupancy proof is retuned identically. The only pre-reduction empirical process is Lemma B.9; its entropy cost is $\log(e/\phi(cr))$. The displayed master condition implies $\phi(cr) \ge C_A \log n/(snr^2) \ge C_A \log n/n$, so this entropy is at most $C \log n$ in the nonvacuous regime. Consequently all discovery, certificate, and refinement deviations are at most a fixed fraction of their order-r population gaps.

Choose the discovery block as the smallest dyadic size satisfying the corresponding polynomial-confidence condition $sn_1\phi(cr)r^2 \ge C_A\{\log n_1 + \log(e/\phi(cr))\}$. This can only reduce the working set relative to the logarithmic-squared choice used in the stated runtime bound. All deterministic candidate, net, and refinement lemmas are unchanged. Union-bounding the retuned n^{-A-12} events over the polynomially many surviving seeds, targets, scales, and constant block assignments gives total failure at most n^{-A}. □

The notation $n^{-\omega(1)}$ is used only after a logarithmic-squared exponent has been established. To see the accounting explicitly, fix any desired polynomial order $A > 0$. The master scale condition is stated with a universal constant that can be enlarged so that each concentration exponent is at least $(A + 10) \log n$. The actual bounds used in the proof are stronger, typically $c \log^2 n$, and hence dominate every fixed multiple of $\log n$.

At discovery, the occupancy failure probability is bounded by

$$\phi(c\rho)^{-1} \exp\{-cn_1\phi(c\rho)\}.$$

The working-set condition implies $n_1\phi(c\rho) \geq C \log^2 n_1$, while the net-cardinality factor contributes only $\log(1/\phi(c\rho))$. Uniform pair concentration spends the larger entropy $O(n_1\phi(c\rho) \log(e/\phi(c\rho)))$, which is already included in Lemma B.9. Thus discovery fails with probability at most n_1^{-A-10} after constants are chosen.

At first refinement, a fixed seed has an exceptional-count failure bounded by $e^{-c \log^2 n}$. The retained family has at most $n_1 \leq n$, and there are only constantly many block assignments. Hence the union bound is at most

$$Cne^{-c \log^2 n} \leq n^{-A-5}$$

for large n. At exact refinement, uniform navigation over all output vertices has conditional failure at most n^{-A-10} per seed, so the same union bound applies. The certificate construction also ranges over at most n coarse probes and n fresh vertices. Its n^2 union factor is absorbed by the same logarithmic-squared exponent.

Finally, conditional estimates are converted into unconditional ones by the tower property. If $\mathcal{F}$ contains the adaptive past and $\Pr(E^c \mid \mathcal{F}) \leq \delta$ almost surely, then

$$\Pr(E^c) = \mathbb{E}[\Pr(E^c \mid \mathcal{F})] \leq \delta.$$

No independence between different successful stages is asserted or needed. A union bound over their unconditional failure probabilities gives total failure at most n^{-A}. Because A was arbitrary, this is precisely the meaning of $1 - n^{-\omega(1)}$ in the theorem.

I.5 Admissible radius audit

Every invocation of the lower-mass condition uses a radius equal to a fixed multiple of r, r_0, ρ, or ρ_0. The theorem's phrase "inside the admissible local range" means that the largest of these multiples is at most r_μ. Similarly, each use of the two-sided link slope is preceded by either an explicit radius inequality or the fuzzy certificate. The constant hierarchy in Lemma A.3 first fixes all multiplicative losses and then restricts r so that their maximum is below r_p. Monotonicity alone is used when a comparison lies farther away; the proof never silently applies the lower slope beyond r_p.

This distinction is important for rough metric spaces. The theorem needs lower volume only at the scales actually sampled and link identifiability only at the scales actually inverted. Behavior of μ or p outside those windows can be arbitrary, except for boundedness and monotonicity needed to control the contribution of outliers.

I.6 Where each model assumption enters

The lower-mass condition is used in four places: net cardinality, uniform occupancy, existence of genuine ball candidates, and the size lower bound for refined clusters. Monotonicity of p is used to compare internal averages with $p(0)$ and to order near and far score levels. The upper local slope controls averaging bias; the lower local slope converts a score drop into metric separation. Conditional independence of edges supplies concentration after sample splitting. The triangle inequality is the only geometric axiom used in cluster construction. Geodesicity appears later, and only for global propagation of locally inverted distances.

This audit also identifies what is not assumed. No covering upper bound, doubling inequality, chart, tangent approximation, curvature bound, or dimension enters the candidate or refinement arguments. The net-cardinality bound follows from disjoint lower-mass balls, not from a doubling hypothesis. Accordingly, constants depend on the link and noise parameters but not on an ambient dimension.

J FROM CLUSTERS TO DISTANCES

Lemma J.1 (Cluster-to-link error)**.** *Suppose U_v, U_w satisfy*

$$X_{U_v} \subseteq B(X_v, r_{\mathrm{cl}}), \qquad X_{U_w} \subseteq B(X_w, r_{\mathrm{cl}}), \qquad |U_v|, |U_w| \geq m.$$

On $\mathcal{E}_{\mathrm{avg}}(V, \lambda, m)$,

$$|N(U_v, U_w) - p(d(X_v, X_w))| \leq \lambda + 2L_p r_{\mathrm{cl}} \tag{72}$$

whenever $d(X_v, X_w) + 2r_{\mathrm{cl}} \leq r_p$. If p is globally L_p-Lipschitz, the condition on the distance is unnecessary.

Proof. For every off-diagonal $(u, u') \in U_v \times U_w$,

$$|d(X_u, X_{u'}) - d(X_v, X_w)| \leq d(X_u, X_v) + d(X_{u'}, X_w) \leq 2r_{\mathrm{cl}}.$$

The Lipschitz bound therefore gives $|p(U_v, U_w) - p(d(X_v, X_w))| \leq 2L_p r_{\mathrm{cl}}$. The uniform pair-average event contributes λ; combining the two terms proves (72). □

Lemma J.2 (Projection and inverse-link stability)**.** *Let $I = [a, b]$, let $p : I \to \mathbb{R}$ be monotone and satisfy the lower slope bound in* (14)*, and let $q = p(t)$ for some $t \in I$. For any $z \in \mathbb{R}$, define*

$$\widehat{t} := p^{-1}\big(\Pi_{p(I)} z\big)\,,$$

where $\Pi_{p(I)}$ is Euclidean projection onto the interval $p(I)$. Then

$$|\widehat{t} - t| \leq \frac{|z - q|}{\ell_p}. \tag{73}$$

Proof. Because $q \in p(I)$, metric projection onto the closed interval $p(I)$ gives $|\Pi_{p(I)} z - q| \leq |z - q|$. The lower slope inequality implies that the inverse of p on its image is $1/\ell_p$-Lipschitz. Applying it to $\Pi_{p(I)} z$ and $q = p(t)$ proves (73). □

Projection is not cosmetic. A noisy average can fall slightly outside the range of the link, in which case the algebraic inverse is undefined. Projection makes the estimator total while preserving the error bound. It also handles endpoint distances without requiring the noise to have a one-sided sign.

Proposition J.3 (Uniform known-link recovery from exact clusters)**.** *Assume that for every v there is a set U_v satisfying*

$$X_{U_v} \subseteq B(X_v, r_{\mathrm{cl}}), \qquad |U_v| \geq m,$$

and that p is globally bi-Lipschitz. On $\mathcal{E}_{\mathrm{avg}}(V, \lambda, m)$, the estimator

$$\widehat{d}(v, w) = p^{-1}\big(\Pi_{p([0, \mathrm{diam}(\mathcal{M})])} N(U_v, U_w)\big)$$

satisfies

$$\max_{v,w} |\widehat{d}(v, w) - d(X_v, X_w)| \leq \frac{\lambda + 2L_p r_{\mathrm{cl}}}{\ell_p}. \tag{74}$$

Proof. Lemma J.1 gives a uniform link-value error of $\lambda + 2L_p r_{\mathrm{cl}}$. Apply Lemma J.2 with $I = [0, \mathrm{diam}(\mathcal{M})]$ to every pair. Since the pair-average event is already uniform over all sets of size at least m, no further union bound over (v, w) is needed. □

This proposition makes the division of labor explicit. Cluster extraction contributes r_{cl}, pair averaging contributes λ, and link inversion contributes the fixed factor $1/\ell_p$. Choosing both first-order terms of size r yields the claimed $O(r)$ rate. Enlarging clusters would improve concentration but worsen geometric bias, whereas shrinking them would do the reverse; the lower-mass profile determines whether a radius-r cluster is large enough to make both terms simultaneously $O(r)$.

Lemma J.4 (Certified local pairs)**.** *Assume the setting of Lemma J.1 and write* $\varepsilon_{\mathrm{pair}} = \lambda + 2L_p r_{\mathrm{cl}}$. *Suppose* $2r_{\mathrm{cl}} + 2\varepsilon_{\mathrm{pair}}/\ell_p \le r_p/4$. *Define*

$$\mathcal{P}_{\mathrm{loc}} := \{(v, w) : N(U_v, U_w) \ge p(r_p/2) - \varepsilon_{\mathrm{pair}}\}. \tag{75}$$

Then every pair with $d(X_v, X_w) \le r_p/2$ *belongs to* $\mathcal{P}_{\mathrm{loc}}$, *while every retained pair satisfies*

$$d(X_v, X_w) \le r_p/2 + 2\varepsilon_{\mathrm{pair}}/\ell_p.$$

For each retained pair, projection to $p([0, r_p])$ *followed by inversion estimates its distance with error at most* $\varepsilon_{\mathrm{pair}}/\ell_p$.

Proof. If $d(X_v, X_w) \le r_p/2$, the range assumption puts every cluster-to-cluster distance inside $[0, r_p]$, and Lemma J.1 gives

$$N(U_v, U_w) \ge p(d(X_v, X_w)) - \varepsilon_{\mathrm{pair}} \ge p(r_p/2) - \varepsilon_{\mathrm{pair}}.$$

Thus the pair is retained.

Conversely, suppose a retained pair had distance larger than $r_p/2 + 2\varepsilon_{\mathrm{pair}}/\ell_p$. Every off-diagonal cluster pair would then have distance at least

$$r_p/2 + 2\varepsilon_{\mathrm{pair}}/\ell_p - 2r_{\mathrm{cl}},$$

which remains in the local window by assumption. Monotonicity, the inverse local slope, and pair concentration would force $N(U_v, U_w) < p(r_p/2) - \varepsilon_{\mathrm{pair}}$, a contradiction. The distance upper bound guarantees that all required comparisons lie in $[0, r_p]$; Lemma J.1 and the $1/\ell_p$-Lipschitz inverse then prove the estimation bound. □

Definition J.5 (Chain property)**.** A finite metric space $(\mathcal{X}, d)$ has the $(r_\circ, \eta)$-chain property if for every $x, y \in \mathcal{X}$ there is a chain $x = x_0, x_1, \ldots, x_k = y$ such that

$$d(x_i, x_{i+1}) \le r_\circ, \qquad \sum_{i=0}^{k-1} d(x_i, x_{i+1}) \le d(x, y) + \eta.$$

Lemma J.6 (Shortest-path extension)**.** *Suppose* $(\mathcal{X}, d)$ *has the* $(r_\circ, \eta)$*-chain property. Assume that all pairs of distance at most* $r_\circ$ *are retained and have estimates* $\widetilde{d}(x, y)$ *satisfying* $|\widetilde{d}(x, y) - d(x, y)| \le \varepsilon$. *Give each retained edge weight* $\widetilde{d}(x, y) + \varepsilon$ *and let* d_{sp} *be the resulting shortest-path metric. Then*

$$d(x, y) \le d_{\mathrm{sp}}(x, y) \le d(x, y) + \eta + C\left(1 + \frac{\mathrm{diam}(\mathcal{X})}{r_\circ}\right)\varepsilon. \tag{76}$$

Proof. Every edge weight is at least its true distance, so the triangle inequality gives the lower bound for every path. For the upper bound, start from a chain in Definition J.5. Remove repeated points and merge consecutive short segments whenever their combined true length is still at most $r_\circ$. Except possibly for one terminal segment, every resulting pair of consecutive edges has combined length greater than $r_\circ$; hence the number of edges is at most $1 + 2\,\mathrm{diam}(\mathcal{X})/r_\circ$ after absorbing the chain excess into the constant. Along this retained chain,

$$\sum_i (\widetilde{d}(x_i, x_{i+1}) + \varepsilon) \le \sum_i d(x_i, x_{i+1}) + 2k\varepsilon.$$

Use the chain-length bound and the estimate for k to obtain (76). □

Lemma J.7 (Sampled chains in a geodesic space)**.** *If* $(\mathcal{M}, d)$ *is geodesic and the sample is an* h*-net of* $\mathcal{M}$, *then the sampled metric has an* $(r_\circ, \eta)$*-chain property whenever* $h < r_\circ/4$, *with*

$$\eta \le C\left(1 + \frac{\mathrm{diam}(\mathcal{M})}{r_\circ}\right)h.$$

Proof. Subdivide a minimizing geodesic from X_v to X_w into segments of length between $r_\circ/2$ and $3r_\circ/4$, except possibly the last. Replace every subdivision point by a sampled point within h. Consecutive sampled points remain at distance below $r_\circ$, and each replacement increases the accumulated length by at most $2h$. The number of subdivisions is $O(1 + \mathrm{diam}(\mathcal{M})/r_\circ)$, proving the claim. □

Lemma J.8 (Explicit sampled-chain count)**.** *Under the assumptions of Lemma J.7, two sampled points at latent distance D can be connected by at most*

$$k \le 2 + \frac{2D}{r_\circ}$$

retained local edges, and the sum of their true lengths is at most

$$D + 2h\left(2 + \frac{2D}{r_\circ}\right). \tag{77}$$

Proof. Let $k_0 = \lceil 2D/r_\circ \rceil$ and divide a minimizing geodesic into k_0 equal pieces. Their lengths are at most $r_\circ/2$. Keep the sampled endpoints and replace each interior division point by a sample within h. A consecutive pair is at distance at most $r_\circ/2 + 2h < r_\circ$, so every edge is retained. The number of edges is $k_0 \le 1 + 2D/r_\circ$; allowing a harmless terminal convention gives the displayed upper bound. Each interior replacement can increase the two incident edge lengths by total at most $2h$, and bounding by $2hk$ yields (77). □

Combining Lemmas J.6 and J.8 gives a transparent accumulated-error formula. If every local edge has estimation error at most ε, then

$$0 \le d_{\mathrm{sp}}(v,w) - d(X_v, X_w) \le C\left(1 + \frac{d(X_v, X_w)}{r_\circ}\right)(h + \varepsilon).$$

The global error is therefore linear in the local error and in the sampling mesh, with a factor equal to the number of link-window-sized segments needed to traverse the space. Geodesicity is used only to supply a chain with this controlled excess.

Lemma J.9 (Global dilation is observationally invariant)**.** *Let $D > 0$, define $d_D(x,y) = Dd(x,y)$, and define $p_D(t) = p(t/D)$. Then for every latent sample and every pair u, v,*

$$p_D(d_D(X_u, X_v)) = p(d(X_u, X_v)).$$

Consequently the entire observed-data law is unchanged under the simultaneous transformation $(d,p) \mapsto (d_D, p_D)$.

Proof. The displayed equality is immediate from the definitions. Conditional means of all pair observations are identical, and the missingness and centered-noise laws are unchanged. Conditional independence then makes the joint laws equal; integrating over the latent sample preserves equality. □

Thus no estimator can identify absolute scale when the link is unknown over a class closed under this transformation. The dilation in Corollary 3.2 is not introduced by the bisection analysis: it is the exact equivalence relation of the observation model. The purpose of the comparison construction is to show that, modulo this equivalence, all remaining pairwise distances can still be aligned consistently.

Proof of Corollary 3.2. Apply Theorem 3.1 and Lemma B.9 with $m = cn\phi(cr)$ and $\lambda = cr$. Equation (7) makes the two events simultaneous with probability $1 - o(1)$.

If p is globally bi-Lipschitz, project $N(U_v, U_w)$ onto the range of p and define

$$\widehat{d}(v,w) = p^{-1}\big(\Pi_{p([0,\mathrm{diam}(\mathcal{M})])} N(U_v, U_w)\big).$$

Projection does not increase link error, and the inverse is $1/\ell_p$-Lipschitz. Lemma J.1 gives $|\widehat{d}(v,w) - d(X_v, X_w)| \le Cr$ uniformly.

If the inverse is available only on $[0, r_p]$, use the threshold in (75). Lemma J.4 implies that all latent pairs of length at most $r_p/2$ are retained, every retained pair has length at most $r_p/2 + Cr$, and its inverted estimate has error at most Cr. These local estimates define a weighted graph. Adding Cr to every edge weight makes each edge an upper bound on the true distance. Lemma J.6, with $r_\circ = r_p/2$, gives

$$d(X_v, X_w) \le \widehat{d}_{\mathrm{sp}}(v,w) \le d(X_v, X_w) + \eta + C\frac{\mathrm{diam}(\mathcal{M})}{r_p} r.$$

For geodesic $\mathcal{M}$, the occupancy event implies that the sample is an $O(r)$-net. Lemma J.7 therefore supplies the required sampled chain with $\eta = O((1 + \operatorname{diam}(\mathcal{M})/r_p)r)$, proving the second part of the corollary.

When p is unknown, the observations are invariant under $d \mapsto Dd$ and $p(t) \mapsto p(t/D)$. The cluster construction supplies the local comparison oracle required by a multiscale midpoint-bisection argument. Applying this argument to the radius-controlled clusters, with $O(r)$ comparison error at each of $O(\log(1/r))$ bisection levels, yields a single $D > 0$ for which

$$\sup_{v,w} |D\widehat{d}(v, w) - d(X_v, X_w)| \le Cr \log(1/r).$$

Only the global dilation is left unidentified, as asserted. □

Proof of Corollary 3.3. If r, s, and $\phi(cr)$ are bounded below, the leading exponential factor in Theorem 3.1 is constant. The working set can be chosen with $n_1 \asymp \log \log n$, while (55) permits $n_2 \asymp \log^2 n$. Candidate enumeration is therefore polylogarithmic, and each surviving candidate makes only $O(n \operatorname{polylog} n)$ edge queries and arithmetic operations. Pair-average evaluation and global inversion have the same near-linear dependence on n. □

K Folding lower bound

K.1 A testing inequality for random targets

Lemma K.1 (Lifted testing-to-risk principle)**.** *Let $X \sim \Pi$, let $Q_0(\cdot \mid X)$ and $Q_1(\cdot \mid X)$ be two conditional observation laws, and let $\Theta_0(X), \Theta_1(X)$ take values in a space equipped with loss metric L. Write $\overline{Q}_j$ for the joint law of (X, Y_j) under $X \sim \Pi$ and $Y_j \mid X \sim Q_j(\cdot \mid X)$. If an event $H \in \sigma(X)$ satisfies $\Pi(H^c) \le \zeta$ and $L(\Theta_0(X), \Theta_1(X)) \ge 2a$ on H, then every observation-measurable estimator $\widehat{\Theta}$ satisfies*

$$\overline{Q}_0\{L(\widehat{\Theta}(Y), \Theta_0(X)) \ge a\} + \overline{Q}_1\{L(\widehat{\Theta}(Y), \Theta_1(X)) \ge a\} \ge 1 - \zeta - \mathrm{TV}(\overline{Q}_0, \overline{Q}_1). \tag{78}$$

Proof. Let E_0 be the event that the first loss is smaller than a. On $H \cap E_0$, the triangle inequality forces $L(\widehat{\Theta}(Y), \Theta_1(X)) \ge a$. Therefore the second error probability is at least $\overline{Q}_1(H \cap E_0) \ge \overline{Q}_0(H \cap E_0) - \mathrm{TV}(\overline{Q}_0, \overline{Q}_1)$. Adding $\overline{Q}_0(H \cap E_0^c)$ and using $\overline{Q}_0(H) = \Pi(H)$ proves (78). □

The lifting by X is essential. The latent distance matrix is random, and the pair of indices witnessing separation is selected from the latent sample. Comparing the joint laws of the common preimage sample and the observations retains that witness while still restricting the estimator to the observed graph.

Lemma K.2 (Expected number of affected cross-pairs)**.** *Let N_1, N_2 be the sample counts in disjoint sets of equal mass m. Then*

$$\mathbb{E}[N_1 N_2] = n(n-1)m^2,$$

and for every $t > 0$,

$$\Pr\{N_1 N_2 > t\} \le \frac{n(n-1)m^2}{t}. \tag{79}$$

Proof. Write $N_1 N_2 = \sum_{i \ne j} \mathbf{1}\{X_i \in A_1, X_j \in A_2\}$. Each ordered summand has expectation m^2, and there are $n(n-1)$ of them. The tail estimate is Markov's inequality. □

The product m^2 is essential. Observations involving exactly one folded cap and the common exterior do not distinguish the models because paired cap points have identical exterior distances. Only pairs with one endpoint in each copy are affected. Thus the effective number of informative comparisons is quadratic in the cap mass even though the number of sampled cap vertices is linear in that mass.

Lemma K.3 (KL bound for a folded experiment)**.** *Let $(\mathcal{X}, d, \mu)$ contain disjoint sets A_1, A_2 exchanged by a measure-preserving involutive isometry T that fixes the complement. Let L fold A_2 onto A_1 and fix all other points, and let $\nu = L_{\#}\mu$. Put*

$$m = \mu(A_1) = \mu(A_2), \qquad \delta_T = \sup_{z \in A_1} d(z, Tz).$$

Assume the Bernoulli link is L_p-Lipschitz and satisfies $p(t) \in [p_-, p_+]$ on the bounded distance range, where $0 < p_- \le p_+ < 1$. Draw a common preimage sample $X_1, \dots, X_n$ independently from μ. Let $\overline{P}_0$ be the joint law of this sample and conditionally independent edges with means $sp(d(X_i, X_j))$, and let $\overline{P}_L$ replace these means by $sp(d(LX_i, LX_j))$. Then

$$D_{\mathrm{KL}}(\overline{P}_0 \| \overline{P}_L) \le C_{p_-,p_+} s L_p^2 \delta_T^2 n(n-1)m^2, \qquad \mathrm{TV}(\overline{P}_0, \overline{P}_L) \le C_{p_-,p_+} \sqrt{s}\, L_p \delta_T nm. \tag{80}$$

The observation marginal of $\overline{P}_L$ is the graph law obtained from n independent samples from ν.

Proof. The two joint laws have the same X-marginal, so the KL chain rule reduces their divergence to the expected conditional edge divergence. For $a, b \in (0, 1)$,

$$D_{\mathrm{KL}}(\mathrm{Ber}(a) \| \mathrm{Ber}(b)) \le \frac{(a-b)^2}{b(1-b)}.$$

Here $a = sp(d(X_i, X_j))$ and $b = sp(d(LX_i, LX_j))$. Since $p \in [p_-, p_+]$ and $s \le 1$, the denominator is at least $sp_-(1-p_+)$, while Lipschitzness bounds the numerator by $s^2 L_p^2 |d(X_i, X_j) - d(LX_i, LX_j)|^2$. Hence an affected edge contributes at most $C_{p_-,p_+} s L_p^2 \delta_T^2$.

Distances are unchanged except when one endpoint lies in A_1 and the other in A_2. Conditional on X, there are $N_1 N_2$ such unordered pairs. Tensorization, followed by Lemma K.2, gives the KL bound. Pinsker's inequality and $n(n-1) \le n^2$ give the TV bound. Finally, the variables $L(X_i)$ are independent with common law $L_{\#}\mu = \nu$, proving the marginal assertion. □

The square in the Bernoulli KL bound is the improvement over a common-uniform edge coupling. Coupling and union-bounding would charge $O(sr)$ per affected pair and require $sn^2\phi(r)^2 r = o(1)$. The lifted KL calculation charges $O(sr^2)$ and only requires $sn^2\phi(r)^2 r^2 = o(1)$, equivalently $\sqrt{s}\, n\phi(r) r = o(1)$.

Lemma K.4 (Separation of the random metric targets)**.** *In the star construction below, couple latent samples by $Y_i = L(X_i)$ and let*

$$D_0(i,j) = d(X_i, X_j), \qquad D_k(i,j) = d(Y_i, Y_j).$$

Let H_k be the event that at least one sampled point lies at each of a_k^+ and a_k^-. Then

$$\|D_0 - D_k\|_\infty \ge 2r_k \qquad \text{on } H_k, \tag{81}$$

and

$$\Pr(H_k^c) \le 2e^{-n\phi(r_k)}. \tag{82}$$

Consequently, if the lifted joint laws in the two experiments have total variation at most δ_k, every estimator $\widehat{D}$ satisfies

$$\begin{aligned} \Pr_0\{\|\widehat{D} - D_0\|_\infty \ge r_k\} + \Pr_k\{\|\widehat{D} - D_k\|_\infty \ge r_k\} \\ \ge 1 - \delta_k - 2e^{-n\phi(r_k)}. \end{aligned} \tag{83}$$

Proof. On H_k, choose indices i^+ and i^- at the two leaves. Their original distance is $D_0(i^+, i^-) = 2r_k$. Folding maps both latent points to a_k^+, so $D_k(i^+, i^-) = 0$. This proves (81). Each leaf has mass $m_k = \phi(r_k)$. A union bound over the two events that a leaf is missed gives

$$\Pr(H_k^c) \le 2(1-m_k)^n \le 2e^{-nm_k},$$

which is (82). Apply Lemma K.1 with joint-law total variation δ_k, target loss $L(D, D') = \|D - D'\|_\infty$, separation $2r_k$, and occupancy failure $2e^{-nm_k}$. The resulting inequality is (83). □

This lemma avoids conditioning the statistical experiment on a particular pair of sampled labels. The distinguishing pair is itself random, but the loss is uniform over the matrix and therefore detects whichever cross-cap pair appears. The estimator receives only the graph labels, not the latent leaf labels; selecting i^+, i^- is an analysis device used to prove that the two random targets cannot both lie inside the same sup-norm ball.

Proof of Theorem 3.5. Because $\phi(r) \to 0$, first restrict the regularity radius, if necessary, so that $\phi(r) \le 1/2$ throughout the range under consideration. Choose $r_k \downarrow 0$ so rapidly that

$$2\sum_{k\ge 1} \phi(r_k) < \frac{1}{2}, \qquad r_{k+1} \le \frac{1}{4} r_k, \tag{84}$$

and construct the countable star metric

$$\mathcal{X}_0 = \{o\} \cup \{a_k^+, a_k^- : k \ge 1\}, \qquad d(o, a_k^\pm) = r_k, \qquad d(a_k^\sigma, a_\ell^\tau) = r_k + r_\ell \quad \text{for distinct leaves.}$$

Assign mass $\phi(r_k)$ to each leaf and the remaining mass to o. This space is lower-ϕ-regular: below r_k the atom $a_k^\pm$ itself supplies mass $\phi(r_k) \ge \phi(r)$, while above r_k the ball reaches the center, whose mass is at least $1/2$.

For completeness, the metric and measure claims require no hidden Euclidean realization. The triangle inequality follows because the unique path between two distinct leaves passes through o. A Cauchy sequence that visits infinitely many leaves must have leaf radii tending to zero by (84), and therefore converges to o; hence $\mathcal{X}_0$ is complete. For every $\varepsilon > 0$, only finitely many leaves have radius exceeding ε, while all remaining leaves lie in $B(o, \varepsilon)$, so the space is totally bounded and thus compact. The masses sum to one by (84).

We also verify the lower-mass condition pointwise. If $x = a_k^\pm$ and $0 < t < r_k$, then $B(x, t)$ contains its atom and

$$\mu(B(x, t)) \ge \phi(r_k) \ge \phi(t).$$

If $t \ge r_k$, the ball contains o and therefore has mass at least $1/2 \ge \phi(t)$. At $x = o$, every ball contains the center atom, again of mass at least $1/2$. Thus the claimed profile holds on the entire selected regularity range, including at the accumulation point.

For each k, fold a_k^- onto a_k^+ and denote the pushforward space by $\mathcal{X}_k$. The involution exchanging these two leaves has displacement $2r_k$, and each folded cap has mass $\phi(r_k)$. Lemma K.3 therefore gives

$$D_{\mathrm{KL}}(\overline{P}_{0,n} \| \overline{P}_{k,n}) \le C s n^2 \phi(r_k)^2 r_k^2, \qquad \mathrm{TV}(\overline{P}_{0,n}, \overline{P}_{k,n}) \le C\sqrt{s}\, n \phi(r_k) r_k. \tag{85}$$

Folding deletes one support point and doubles the mass of its mate. It cannot decrease the mass of any ball centered at a surviving point: balls not meeting the folded leaf are unchanged, while balls meeting it receive its mass at the paired leaf no farther from the center by the star symmetry. Hence every $\mathcal{X}_k$ remains lower-ϕ-regular.

From KL indistinguishability to estimation error. Choose a diagonal sequence $k = k(n)$ such that

$$n\phi(r_{k(n)}) \longrightarrow \infty, \qquad \sqrt{s_n}\, n \phi(r_{k(n)}) r_{k(n)} \longrightarrow 0. \tag{86}$$

The freedom to thin the sequence (r_k) in (84) permits this diagonal choice whenever the displayed indistinguishability regime of the theorem is considered. Using (85) gives $\mathrm{TV}(\overline{P}_{0,n}, \overline{P}_{k(n),n}) = o(1)$.

Let H_k be the event that the sample from $\mathcal{X}_0$ contains at least one vertex at each of a_k^+ and a_k^-. If $m_k = \phi(r_k)$, then

$$\Pr_0(H_k^c) \le \Pr\{N_k^+ = 0\} + \Pr\{N_k^- = 0\} \le 2(1 - m_k)^n \le 2e^{-n m_k} = o(1). \tag{87}$$

On H_k, choose the first sampled vertex in each leaf and call their indices i_k^+ and i_k^-. In $\mathcal{X}_0$ their latent distance is $2r_k$; after folding, the corresponding two latent points coincide and their distance is zero. No real number can be within $r_k/2$ of both values. Therefore, for every graph-measurable distance estimator $\widehat{D}$, the standard two-point testing inequality gives

$$\begin{aligned} \Pr_0\Big\{\|\widehat{D} - D_0\|_\infty \ge r_k/2\Big\} + \Pr_k\Big\{\|\widehat{D} - D_k\|_\infty \ge r_k/2\Big\} \\ \ge 1 - \mathrm{TV}(\overline{P}_{0,n}, \overline{P}_{k,n}) - \Pr_0(H_k^c) = 1 - o(1). \end{aligned} \tag{88}$$

Thus at least one of the two lower-ϕ-regular models has nonvanishing probability of uniform error at least $r_k/2$, ruling out uniform $o(r_k)$ recovery.

Finally, take $p(t) = b - at$ on the bounded distance range, with $a > 0$ small enough that p stays inside a fixed interval $[p_-, p_+] \subset (0, 1)$, and truncate it outside that range if desired. This link is monotone and bi-Lipschitz with equal local slopes. The construction therefore uses no flat part of the link. The obstruction is combinatorial: only cross-pairs between two sets of mass $\phi(r_k)$ reveal the fold, giving $n^2\phi(r_k)^2$ potentially informative pairs. Each contributes $O(sr_k^2)$ KL divergence, which yields precisely the aggregate scale in (86) after Pinsker's inequality. □

L SUPPLEMENTARY EXPERIMENTAL DETAILS

The experiments test the two terms in Equation (6) and the role of fresh-edge refinement, rather than approximate the exhaustive worst-case search used in the proof. For each target v, the rough proxy ranks vertices by a relation-profile score on the discovery block. The refined proxy adds the same target-centered score computed from the independent refinement block. Both stages therefore remain centered at v; refinement does not replace the target by the density center of its rough candidate. The oracle proxy contains the $\lceil\sqrt{n}\rceil$ latent nearest neighbors and measures residual cluster-averaging error.

For a target v and cluster U_v, define the evaluation radius by $R_v := \max_{u\in U_v} d(X_u, X_v)$. Each repetition reports the median of R_v over 24 targets, and the tables average this median over repetitions. Thus radius is used only to diagnose geometric bias, never by an estimator. The main experiment uses 20 repetitions. Sensitivity experiments use 10 repetitions at $n = 1200$, varying one parameter at a time around $s = 0.75$ and $\sigma = 0.15$. All random seeds are fixed by the experiment script.

Table 4: Cluster-radius ablation and extraction time for Table 3. Radius entries are mean median target-to-cluster radii; time is the mean wall-clock time for constructing both observable proxies for all 24 targets.

Space	n	Rough radius	Refined radius	Oracle radius	Time (s)
Interval	300	0.147	0.120	0.026	0.007
	600	0.121	0.101	0.019	0.023
	1200	0.102	0.088	0.014	0.117
Metric tree	300	0.562	0.428	0.104	0.007
	600	0.397	0.175	0.077	0.022
	1200	0.151	0.118	0.055	0.131

Refinement reduces radius and MAE at every reported size. The largest radius change occurs on the tree at $n = 600$, from 0.397 to 0.175, while MAE falls from 0.054 to 0.041. At $n = 1200$, refined-to-oracle radii are 0.088 versus 0.014 on the interval and 0.118 versus 0.055 on the tree, so neighborhood construction remains the main error source. On an Apple M1 with Python 3.9.1 and NumPy 2.0.2, construction time for 24 targets grows from about 0.007 seconds at $n = 300$ to 0.12–0.13 seconds at $n = 1200$. These timings describe the proxy implementation, not the worst-case complexity bound.

Table 5: Sensitivity of refined-proxy MAE at $n = 1200$, averaged over 10 repetitions. The observation-rate panel fixes $\sigma = 0.15$; the noise panel fixes $s = 0.75$.

Parameter	Space	Parameter value			
		0.25	0.50	0.75	1.00
Observation rate s	Interval	0.032	0.018	0.015	0.013
	Metric tree	0.082	0.033	0.025	0.024
		0.05	0.15	0.30	
Noise s.d. σ	Interval	0.005	0.015	0.034	
	Metric tree	0.012	0.024	0.062	

The trends agree with Equation (7): more observed edges improve target-centered profiles and cluster averages, while larger noise degrades neighborhood selection and inversion. Refined-proxy coverage is one in every sensitivity setting because a cluster pair contains many independent evaluation edges; the direct baseline is defined only when the single target pair lies in the evaluation block.