DISCRETIZATION-FREE EXACT RECOVERY IN GEOMETRIC COMMUNITY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Many graph-learning problems combine relational observations with continuous geometric side information, yet existing exact-recovery methods for geometric community detection propagate labels according to a fixed spatial discretization. We instead introduce an information-adaptive algorithm for the Geometric Hidden Community Model that operates directly on the observed geometry: at each step it labels the frontier vertex whose already labeled neighbors provide the strongest worst-case Hellinger evidence, so the propagation order is determined by statistical information rather than by a spatial mesh. This perspective also permits heterogeneous information across communities and distances: different community pairs may rely on different witness communities and different distance ranges. In particular, our condition covers canonical symmetric multi-community models, including the geometric stochastic block model, for which the previous all-witness condition fails whenever there are at least three communities. Under an explicit witness-connectivity condition and the known Chernoff–Hellinger information condition, we prove polynomial-time exact recovery; in the standard fully distinguishable regime, the sufficient boundary coincides with the information-theoretic threshold. Experiments show that information-adaptive ordering substantially reduces propagation errors relative to spatial and degree-based schedules, succeeds beyond the regime covered by our proof, and performs competitively with calibrated block propagation on a canonical geometric block-model benchmark.

1 INTRODUCTION

Many graph-learning problems provide more structure than the graph alone: nodes carry observed locations or continuous embeddings, and interaction distributions depend jointly on latent structure and geometric separation. Community detection is a clean instance: the geometry is observed, while the community labels are latent. A natural principle is therefore to exploit the geometry at its native resolution and propagate labels where the available observations are statistically most informative.

We study this question in the Geometric Hidden Community Model (GHCM) introduced by (Gaudio et al., 2025). Vertices form a Poisson point process in a continuous spatial domain, their locations are observed, and edge-weight distributions depend jointly on the latent labels and geometric distance. Exact community recovery is thus a structured prediction problem on a graph with continuous geometric side information. Recent work established information-theoretic limits and efficient recovery procedures for this model and related geometric block models (Gaudio & Jin, 2025; Gaudio et al., 2025; Gaudio & Jin, 2026). Their propagation mechanisms, however, first discretize space into blocks and let this auxiliary structure determine how labels spread. This raises a basic question: should propagation be organized by a fixed geometric scaffold, or by the statistical information that the geometry makes available?

The issue: information need not be uniform across communities or distances. Existing propagation guarantees for the GHCM require any two communities $a \neq b$ to be separated by *every* third community w, that is, $P_{aw} \neq P_{bw}$ (Gaudio & Jin, 2026, Assumption 3). This fails even in the canonical symmetric setting. Suppose $P_{aa} = P_{\text{in}}$ for every a and $P_{ab} = P_{\text{out}}$ for every $a \neq b$, as in the three-community GSBM with $P_{\text{in}} = \text{Ber}(0.8)$ and $P_{\text{out}} = \text{Ber}(0.2)$. For any $a \neq b$ and $w \notin \{a, b\}$,

$$P_{aw} \;\equiv\; P_{\text{out}} \;\equiv\; P_{bw},$$

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.

so the third community is blind to the pair at every distance. The condition is therefore vacuous at $k = 2$ and fails for every $k \geq 3$, although the pair itself remains informative. More generally, different pairs may be distinguishable through different communities and on different distance ranges. We benchmark on the symmetric model in Section 5.5.

Our approach: propagate where the information is. We replace block-wise propagation by an adaptive vertex-level rule. Starting from a correctly initialized seed region, the algorithm evaluates every unlabeled frontier vertex through the worst-case Hellinger affinity information supplied by its labeled visible neighbors. It propagates to the vertex with the largest available information, assigns its maximum-likelihood label, and updates the scores. Thus geometry determines which evidence is available, while statistical information determines the propagation order. The resulting method can be viewed as model-based message passing with an adaptive schedule and requires no spatial partition or resolution parameter.

Our main contributions are the following.

- **Information-adaptive propagation.** We introduce a polynomial-time exact-recovery algorithm operating directly on the observed continuous vertex locations. Its propagation schedule is constructed online from the statistical evidence in the labeled neighborhood rather than from a predetermined mesh. The discretization is removed from the algorithm, not from the analysis: our proof still uses a virtual block structure to certify that enough information is available.
- **Recovery from heterogeneous geometric information.** We require only that each pair $a \neq b$ is separated by *some* community on *some* interval, chosen pair by pair, with sufficient witness mass (Assumption 1). Different pairs may therefore use different witnesses and distance ranges. The earlier all-witness condition is unsatisfiable on every symmetric model with $k \geq 3$, including the geometric SBM (Remark 4), whereas ours holds there. Example 5 gives a model covered by Theorem 3 and by no existing result.
- **Sharp guarantees without a spatial resolution parameter.** We prove exact recovery for both the adaptive algorithm and a generalized block procedure under the same witness condition, separating the weaker structural requirement from the removal of discretization. Whenever witness connectivity is automatic, the sufficient recovery boundary coincides with the known Chernoff–Hellinger information-theoretic threshold. The block method, however, requires a resolution whose admissible range exists only under the witness-connectivity inequality (Proposition 7); the adaptive algorithm runs unchanged outside this certified range.
- **Why the adaptive order matters.** Replacing affinity-maximization in Phase II by a spatial ordering, or by a rule maximizing the number of labeled neighbors at informative distances, costs up to an order of magnitude in propagation errors. The informed rules collect essentially the same mean affinity information; they differ in the lower tail, precisely the quantity controlling the one-step error bound. A canonical GSBM benchmark further compares the adaptive rule with calibrated block propagation.

Related work

Community detection in spatially embedded networks was initiated by (Sankararaman & Baccelli, 2018) through the Geometric Stochastic Block Model (GSBM), for which (Gaudio & Jin, 2025) developed an efficient exact-recovery procedure based on local propagation and likelihood refinement, complementing the lower bounds of (Abbe et al., 2021). (Avrachenkov et al., 2024) allowed general distance-dependent edge probabilities in one dimension; (Gaudio et al., 2025) extended this framework to arbitrary joint edge-weight distributions in the GHCM, and Gaudio & Jin (2026) characterized the information-theoretic threshold and gave an efficient algorithm under an all-witness condition. That condition excludes every symmetric model with $k \geq 3$ (Remark 4), and hence Example 5. Soft geometric block models with unknown locations (Avrachenkov et al., 2021) and spherical variants (Galhotra et al., 2018; 2023) have also been studied. Our focus is complementary: rather than fixing a propagation structure on the continuous geometry, we let the statistical information available in that geometry determine the propagation order.

2 MODEL, ALGORITHM AND RECOVERY GUARANTEES

We consider the following GHCM of (Gaudio & Jin, 2026).

Definition 1 (GHCM)**.** Fix a dimension d, an intensity $\lambda > 0$, community labels Z with $|Z| = k$, priors $\pi \in \mathbb{R}^k$ with $\pi_i > 0$, and symmetric edge-weight distributions $P_{ij}(y) = P_{ji}(y)$ indexed by the distance y. We sample $G \sim \mathrm{GHCM}(\lambda, n, r, \pi, P(\cdot), d)$ in three steps. The vertex set V is a Poisson point process of intensity λ on $\mathcal{S}_{d,n} := [-n^{1/d}/2, n^{1/d}/2]^d$, the torus of volume n; each $v \in V$ independently receives a label $\sigma^*(v)$ with $\mathbb{P}(\sigma^*(v) = i) = \pi_i$; and, conditionally on locations and labels, the weights

$$X_{uv} \sim \bar{P}_{\sigma^*(u)\sigma^*(v)}(\|u - v\|), \qquad \bar{P}_{\sigma^*(u)\sigma^*(v)}(y) := P_{\sigma^*(u)\sigma^*(v)}\big(y/(\log n)^{1/d}\big),$$

are drawn independently over pairs $u \neq v$, where $\|\cdot\|$ is the toroidal metric and x_{uv} is the observed value. The rescaling places the expected number of visible neighbors at $\lambda\nu_d r^d \log n$, the logarithmic scale at which exact recovery is possible. Here r is the least distance beyond which all P_{ab} coincide with a common point mass. We call $r(\log n)^{1/d}$ the *visibility radius* and write $u \leftrightarrow v$ when $\|u - v\| \le r(\log n)^{1/d}$. Appendix F.1 gives the measure-theoretic conventions and the formal definition of r.

Symmetries in π and P make the labels identifiable only up to permutations preserving the model parameters. The following definition, adapted from (Gaudio et al., 2025) to distance-dependent kernels, makes this precise. A permutation $\omega : Z \to Z$ is called permissible if $\pi_i = \pi_{\omega(i)}$ and $P_{ij}(y) = P_{\omega(i)\omega(j)}(y)$ for all $i, j \in Z$ and a.e. $y \in [0, r]$. Let $\Omega_{\pi,P}$ denote the set of such permutations.

We say that an estimator $\tilde{\sigma}_n$ achieves *exact recovery* if

$$\mathbb{P}(\exists\, \omega \in \Omega_{\pi,P} : \tilde{\sigma}_n = \omega \circ \sigma_n^*) \longrightarrow 1 \qquad \text{as } n \to \infty.$$

For $t \in (0, 1)$ and probability measures $P, Q \ll \mu$ with densities p, q, respectively, define

$$\phi_t(P, Q) := \int_{\mathbb{X}} p(x)^t q(x)^{1-t}\, \mu(dx). \tag{1}$$

Definition 2 (Hellinger affinity information)**.** For $u \in \mathcal{S}_{d,n}$ let $N(u) := \{v \in \mathcal{S}_{d,n} : u \leftrightarrow v, u \neq v\}$ be the *neighborhood* of u. Let $u \in \mathcal{S}_{d,n}$, $S \subseteq V \cap N(u)$, and let $\hat{\sigma} : S \to Z$ be a labeling. We define

$$\mathcal{I}_S(u) := \min_{a \neq b} \sum_{v \in S} -\log \phi_{1/2}\big(\bar{P}_{a\hat{\sigma}(v)}(\|u - v\|), \bar{P}_{b\hat{\sigma}(v)}(\|u - v\|)\big)\,. \tag{2}$$

Any fixed $t \in (0, 1)$ would serve. What matters is that the exponent is not optimized over t, so that scores update additively, and that $\phi_t(P, Q) < 1$ whenever $P \neq Q$, which holds for every $t \in (0, 1)$. We take $t = 1/2$ for its symmetry in the two arguments.

The affinity score leads directly to our estimator. Starting from a locally initialized seed region $B_{\mathrm{init}} := [-r(\log n)^{1/d}, r(\log n)^{1/d}]^d$, Algorithm A repeatedly labels the frontier vertex with the largest worst-case affinity information, and finally performs a likelihood refinement sweep. The seed is labeled by a maximum a posteriori (MAP) estimator, and each selected vertex receives its maximum-likelihood label given its labeled neighbors in Phase II. Unlike block propagation, Phase II does not require a spatial partition or resolution parameter.

Algorithm A Discretization-free Exact Recovery

Input: $G \sim \mathrm{GHCM}(\lambda, n, r, \pi, P(\cdot), d)$.
Output: An estimated community labeling $\widetilde{\sigma} : V \to Z$.

Phase I (Initialization):
Let $\hat{V} := V \cap B_{\mathrm{init}}$.
Apply `Maximum a Posteriori` (Algorithm 1) on input $(G, \hat{V})$ to obtain a labeling $\hat{\sigma}$ on $\hat{V}$.

Phase II (Propagation):
while $V \setminus \hat{V} \neq \emptyset$ **do**
  if $\{v \in V \setminus \hat{V} : N(v) \cap \hat{V} \neq \emptyset\} = \emptyset$ **then**
    Set $\hat{\sigma}(v) = *$ for all $v \in V \setminus \hat{V}$ and exit the loop.
  Let $u := \arg\max_{v \in V \setminus \hat{V} : N(v) \cap \hat{V} \neq \emptyset} \mathcal{I}_{N(v) \cap \hat{V}}(v)$.
  Apply `Propagate` (Algorithm 2) on input $(G, \hat{V}, \{u\}, \hat{\sigma}_{\hat{V}})$ to obtain a labeling $\hat{\sigma}(u)$.
  $\hat{V} \leftarrow \hat{V} \cup \{u\}$.

Phase III (Refinement):
for $u \in V$ **do**
  Apply `Refine` (Algorithm 3) on input $(G, u, \hat{\sigma})$ to obtain a labeling $\tilde{\sigma}(u)$.

Next we give conditions under which we establish exact recovery. For $a \neq b$ and $I \subseteq [0, r]$, define the set of *witness communities*

$$W_{ab}(I) := \left\{ w \in Z : \inf_{t \in (0,1)} \phi_t\big(P_{aw}(y), P_{bw}(y)\big) < 1 \text{ for a.e. } y \in I \right\}.$$

Thus, $w \in W_{ab}(I)$ precisely when the edge-weight distributions from communities a and b to community w are distinguishable for almost every distance in I.

Assumption 1 (Identifiability of pairwise distributions)**.** Assume that

$$\lambda \nu_d \sup_{0 < \ell \le r} \min_{a \neq b} \sup_{\ell \le s \le r} \ell^d \sum_{w \in W_{ab}([s-\ell, s])} \pi_w > c_d,$$

with

$$c_d := \begin{cases} 1, & d \ge 2, \\ 2, & d = 1. \end{cases}$$

where ν_d denotes the d-dimensional volume of the unit ball in $\mathbb{R}^d$.

Assumption 1 asks for a common length scale ℓ such that every pair of communities is separated by a set of witness communities whose combined prior probabilities are large enough, on some window of width ℓ, to let labels propagate between them. Figure 1 previews what this buys over the all-witness condition of (Gaudio & Jin, 2026): region A is information-theoretically impossible; region C is where Theorem 3 applies; region B is the gap between the two, where the information-theoretic condition holds but our proof technique does not yet certify recovery. We return to all three regions below and again in Section 5.

To quantify the distinguishability between communities, we define the *Chernoff–Hellinger (CH) divergence* (Abbe & Sandon, 2015). For $a \in Z$ and $y \in [0, r]$, let

$$P_a(\,\cdot\,; y) := \big(P_{ac}(\,\cdot\,; y)\big)_{c \in Z}$$

denote the vector of edge-weight distributions from a vertex in community a to vertices in the respective communities at distance y. For communities $a, b \in Z$ we define the CH divergence by

$$D_+(P_a \| P_b) := 1 - \inf_{t \in (0,1)} \int_0^r \frac{d y^{d-1}}{r^d} \sum_{c \in Z} \pi_c \, \phi_t\big(P_{ac}(y), P_{bc}(y)\big) \, dy. \tag{3}$$

Thus, $D_+(P_a \| P_b)$ measures the average distinguishability between communities a and b, where the edge-weight distributions are averaged over the distance distribution and over the community labels.

Define

$$\lambda_{\mathrm{wit}} := c_d\Big(\nu_d \sup_{0<\ell\le r} \min_{a\neq b} \sup_{\ell\le s\le r} \ell^d \sum_{w\in W_{ab}([s-\ell,s])} \pi_w\Big)^{-1}, \qquad \lambda_{\mathrm{IT}} := \Big(\nu_d r^d \min_{a\neq b} D_+(P_a\|P_b)\Big)^{-1}.$$

Then Assumption 1 is precisely $\lambda > \lambda_{\mathrm{wit}}$, while the Chernoff–Hellinger condition in Theorem 3 below is $\lambda > \lambda_{\mathrm{IT}}$. Region B of Figure 1 is therefore $\lambda_{\mathrm{IT}} < \lambda \le \lambda_{\mathrm{wit}}$.

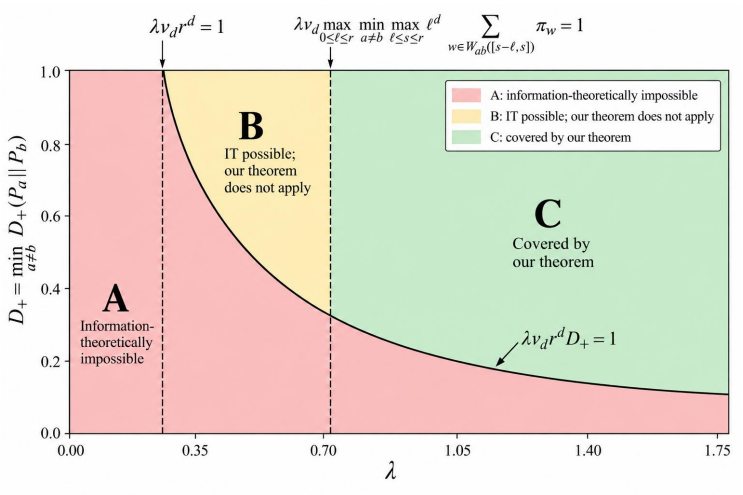


Figure 1: Information-theoretic and witness-connectivity regimes. Region A is information-theoretically impossible. Region C is covered by Theorem 3. Region B is the intermediate regime in which Condition (5) holds but the auxiliary partition used in the proof of Theorem 3 is not guaranteed to percolate, so the theorem does not apply; empirically (Section 5) the adaptive algorithm recovers throughout it.

The next assumption says that at sufficiently short distances, assigning two vertices from the same true community to two different candidate communities produces a uniformly distinguishable pair distribution.

Assumption 2 (Initialization separation)**.** There exist $r_{\mathrm{init}} \in (0, r]$ and $\Phi_{\mathrm{init}} \in (0, 1)$ such that, for every $a, b, c \in Z$ with $b \neq c$,

$$\phi_{1/2}\left(P_{bc}(y), P_{aa}(y)\right) \le \Phi_{\mathrm{init}}$$

for a.e. $y \in [0, r_{\mathrm{init}}]$. Here a is the true community of two nearby vertices and $b \neq c$ are two *candidate* labels assigned to them, so the assumption says that splitting a same-community pair is detectable from the edge between them at short range.

Assumption 3 (Bounded log-likelihood difference)**.** For any communities $a, b, a', b' \in Z$, there exists $\rho > 0$ such that for all $x \in \mathbb{X}$ and $y \in [0, r]$,

$$\left|\log \frac{p_{ab}(x; y)}{p_{a'b'}(x; y)}\right| < \rho. \tag{4}$$

Admissible edge distributions. Assumption 3 holds when $\mathbb{X}$ is finite or countable with $\inf p_{ab}(x; y) > 0$, as for Bernoulli or categorical weights, and more generally when all $P_{ab}(y)$ share a compact support on which their densities are bounded above and away from zero, uniformly in $y \in [0, r]$; in particular it forces mutual absolute continuity, so $\mathrm{Ber}(0)$ against $\mathrm{Ber}(p)$ is excluded.

The following theorem is our main result. Algorithm B is the block-based benchmark (Appendix E).

Theorem 3 (Achievability)**.** *Let Assumptions 1, 2, and 3 hold and suppose that*

$$\lambda\nu_d r^d \min_{a\neq b} D_+(P_a\|P_b) > 1. \tag{5}$$

Then Algorithms A and B achieve exact recovery in polynomial time. Moreover, for fixed k and unit-cost $\phi_{1/2}$ evaluations (Section 3), the propagation phase of Algorithm A runs in $\mathcal{O}(n \log^2 n)$ time with high probability.

Assumption 1 asks only that each pair $a \neq b$ is separated by *some* witness community on *some* interval of a common length ℓ. This replaces the condition that $P_{aw} \neq P_{bw}$ for every w and a.e. $y \in [0, r]$ (Gaudio & Jin, 2026, Assumption 3). The latter is unsatisfiable on the canonical model in this setting, as the following remark shows.

Remark 4 (Pair-blind third communities in symmetric models). Call a GHCM *symmetric* if there are kernels $P_{\text{in}}, P_{\text{out}}$ with $P_{aa}(y) = P_{\text{in}}(y)$ for every a and $P_{ab}(y) = P_{\text{out}}(y)$ for every $a \neq b$. For $k \geq 3$, any $a \neq b$ and any $w \notin \{a, b\}$ satisfy $P_{aw} \equiv P_{\text{out}} \equiv P_{bw}$, so $W_{ab}(I) = \{a, b\}$ for every interval I on which $P_{\text{in}} \neq P_{\text{out}}$, and (Gaudio & Jin, 2026, Assumption 3) fails at every distance and every parameter value. Assumption 1 instead reads

$$\lambda \nu_d \ell^d \min_{a \neq b} (\pi_a + \pi_b) > c_d.$$

The distinctness condition is thus vacuous at $k = 2$ and does not hold for $k \geq 3$.

Example 5. *Let $d = 2$, $r = 3$, $k = 3$ and $\pi = \left(\frac{1}{2}, \frac{3}{10}, \frac{1}{5}\right)$. The kernel is the symmetric kernel of Remark 4 above with $P_{\text{in}} = \text{Ber}(p)$ and $P_{\text{out}} = \text{Ber}(1-p)$, modified so that an intermediate band of distances carries no information. For $p \in (0,1) \setminus \{\frac{1}{2}\}$ and $a, b \in \{1, 2, 3\}$ set*

$$P_{ab}(y) := \begin{cases} \text{Ber}(p), & a = b,\ y \in [0, r/3) \cup (2r/3, r], \\ \text{Ber}(1-p), & a \neq b,\ y \in [0, r/3) \cup (2r/3, r], \\ \text{Ber}(1/2), & y \in [r/3, 2r/3], \end{cases} \tag{6}$$

where Ber(q) denotes the Bernoulli distribution with parameter $q \in [0, 1]$. Observations are thus informative at short and long range, while the intermediate band carries no community information. Theorem 3 applies on the part of Region C with $D_+ \leq \frac{1}{3}$ (the full range Example 5 can reach), which no existing result covers (Appendix B.1).

Condition (5) is known to be necessary for exact recovery (Gaudio & Jin, 2026, Theorem 1). An important special case makes the sharpness explicit.

Corollary 6 (Sharp threshold under full distinguishability)**.** *Suppose $d \geq 2$ and, for every $a \neq b$ and every $w \in Z$,*

$$P_{aw}(y) \neq P_{bw}(y) \qquad \text{for a.e. } y \in [0, r].$$

If Assumptions 2 and 3 hold, then both Algorithms A and B achieve exact recovery whenever Condition (5) *holds. Hence their sufficient recovery boundary matches the known information-theoretic boundary in this regime. Region B of Figure 1 is empty here.*

By Remark 4, this regime is disjoint from the symmetric models: it is the opposite extreme, where every community witnesses every pair. Assumption 1 covers both.

3 Interpretation and comparison

Fixing the witness scale. By Assumption 1, we may fix a common length $\ell \in (0, r]$ and, for every pair $a \neq b$, an interval $I_{ab} \subseteq [0, r]$ with $|I_{ab}| = \ell$ such that

$$\lambda \nu_d \ell^d \sum_{w \in W_{ab}(I_{ab})} \pi_w > c_d \qquad \text{for every } a \neq b. \tag{7}$$

Let

$$\pi_{\text{wit}} := \min_{a \neq b} \sum_{w \in W_{ab}(I_{ab})} \pi_w,$$

so that $\lambda \nu_d \ell^d \pi_{\text{wit}} > c_d$. Thus ℓ is a common informative length scale and π_{wit} is the smallest witness mass available to any community pair at that scale.

Interpretation of Algorithm A. Algorithm A has three phases. First, we initialize on the local seed region B_{init} using the maximum a posteriori (MAP) principle from (Gaudio & Jin, 2026). The seed contains $\lambda(2r)^d \log n + O_{\mathbb{P}}(\sqrt{\log n})$ vertices, where $O_{\mathbb{P}}$ denotes a term bounded in probability and all asymptotic statements are as $n \to \infty$. Thus exhaustive MAP examines $k^{|V_{\text{init}}|} = n^{\lambda(2r)^d \log k + o_{\mathbb{P}}(1)}$ candidate labelings. Its exponent is large, so Section 5.4 substitutes a practical initializer.

The second phase is the departure from block propagation. Suppose a set $\hat{V}$ has already been labeled. For every unlabeled vertex u, $\mathcal{I}_{\hat{V}\cap N(u)}(u)$ measures the least Hellinger information available for distinguishing any two candidate labels using the currently labeled visible neighborhood. We propagate next to the vertex maximizing this quantity and assign its maximum-likelihood label from the already labeled neighbors. Thus the algorithm constructs its own propagation order from the information geometry of the instance; no spatial mesh is used. Because the exponent is fixed, adding one labeled neighbor produces an additive update to each of the $O(k^2)$ pair scores. For fixed k and unit-cost $\phi_{1/2}$ evaluations, there are $O(n \log n)$ visible vertex pairs with high probability and a heap update costs $O(\log n)$, giving $O(n \log^2 n)$ propagation time; the visible pairs are enumerated once in $\Theta(n \log n)$ expected time by bucketing into cells of side $r(\log n)^{1/d}$ (Proposition 33).

Finally, we apply the local refinement of (Gaudio & Jin, 2026) (Algorithm 3), which relabels each $u \in V$ by maximum likelihood over its visible labeled neighbors, with complexity $O(n \log n)$.

Block-based comparison. We also analyze a generalized block procedure, Algorithm B (Appendix E), which separates the witness argument from the effect of removing discretization. Algorithm B partitions space into blocks at a fixed scale χ, certifies blocks containing sufficiently many witnesses, and propagates labels through chains of such certified blocks. When a new block is processed, its vertices are labeled using their visible neighbors in the certified source blocks already reached by the propagation. Unlike Algorithm A, the resulting estimator therefore depends on the spatial resolution χ. Proposition 7 shows that a theoretically admissible choice of χ exists exactly when $\lambda > \lambda_{\mathrm{wit}}$; Algorithm A requires no such parameter.

Proposition 7 (Witness connectivity is exactly the block-scale condition)**.** *Fix $d \geq 1$ and model parameters $(\lambda, r, \pi, P(\cdot))$. The block construction of Appendix E is determined by a common length $\ell \in (0, r]$, witness intervals $\{I_{ab}\}_{a\neq b}$ with $|I_{ab}| = \ell$, and a block scale $\chi > 0$ required to satisfy* (8) *for $d = 1$ and* (9) *for $d \geq 2$. Call such a triple an* admissible block configuration*. Then an admissible block configuration exists for some $(\ell, \{I_{ab}\})$ if and only if $\lambda > \lambda_{\mathrm{wit}}$, i.e. if and only if Assumption 1 holds.*

These are the connectivity requirements on the auxiliary block structure; only their solvability matters here. Proposition 7 separates the algorithms through the free resolution parameter, not through the recovery guarantee. In Region C both are covered by Theorem 3. In Region B Algorithm A is unchanged and uses no spatial parameter, whereas no admissible block configuration exists. Block propagation may still run with a hand-tuned resolution, as in Section 5.5, but the analysis supplies none and the choice matters (Appendix C.3).

What the partition is doing. Assumption 1 enters the analysis in two distinct ways, only one tied to discretization. The qualitative requirement that each pair admit a witness community on some interval ensures pairwise identifiability and is used throughout. The quantitative requirement on the witness mass is used to make an auxiliary cubic grid connected, so that δ-spread blocks form a connected structure (Proposition 24) from which the $\Omega(\log n)$ affinity bound of Lemma 31 follows. For Algorithm A this requirement is a limitation of our proof technique rather than a known algorithmic barrier: Figure 3 (Appendix C.1) finds recovery tracking the information-theoretic boundary throughout Region B.

4 Proof Overview

Initialization. The MAP estimator recovers all labels on B_{init}, up to a permissible relabeling, with probability $1 - o(1)$; the argument combines Assumption 2 with concentration of community counts and likelihood ratios, treating the random Poisson seed size via the Mecke equation.

Propagation is the main difficulty, and both algorithms are handled by a common argument. Assumption 1 guarantees that witness-rich blocks form a connected geometric structure with high probability (Proposition 24). For Algorithm B this directly supplies informative labeled blocks; for Algorithm A the partition enters only the analysis. Lemma 30 produces, at every step before termination, an unlabeled vertex whose labeled witness-rich blocks span the relevant distance ranges; since Algorithm A selects the vertex of maximal affinity information, its choice carries at least as much (Lemma 31). Lemma 25 converts this $\Omega(\log n)$ lower bound into a polynomially small

conditional error probability, Assumption 3 ensuring that boundedly many incorrect neighbors perturb the comparison by $O(1)$. The order is data dependent, but the next vertex is selected before the edge observations used to label it are revealed, so the bound holds conditionally on the history. This is a real constraint on the design. A rule that ordered instead by the most decisive tentative likelihood ratio would inspect the very edges that then label the selected vertex, and the conditional bound would fail. A predictable-thinning and first-failure argument then yields $O(1)$ errors per visibility neighborhood.

Refinement. The preliminary labeling differs from the truth at $o(\log n)$ vertices in every neighborhood, while under (5) the refinement likelihood gap is of order $\log n$. The per-vertex failure probability is $n^{-(1+\Omega(1))}$, and a union bound gives exact recovery.

5 Experiments

5.1 Setup

We study two three-community models specified in Appendix B.1, together with the canonical GSBM of Section 5.5, chosen to isolate the two quantities governing affinity-guided propagation: the witness mass π_{wit} and the information $-\log \phi_{1/2}$ carried by a single witness. Example 5 has abundant, moderately informative witnesses; the long-range instance has rare but highly informative ones, its hardest pair separated only by the rarest community. On both, (Gaudio & Jin, 2026, Assumption 3) fails identically.

Throughout $n = 10^4$, methods run on matched instances with identical seed labels, PROPAGATE, REFINE, and frontiers; only the Phase-II selection rule differs. Besides affinity-max, we use nearest-first, which sweeps the frontier outward from the seed region and is the vertex-level analogue of block propagation; informative-degree-max, which selects the vertex having the largest number of labeled neighbors in distance bands on which the model contains community information; and uniform random and affinity-min as floors. All five rules select from the same frontier, the unlabeled vertices with at least one labeled visible neighbor, and differ only in which frontier vertex they pick. Outcomes are compared with McNemar's exact test.

5.2 Adaptive ordering suppresses propagation errors

Among the informed rules, at $\lambda = 1.2\lambda_{\mathrm{IT}}$, affinity-max makes 1.9 times fewer Phase-II errors than nearest-first on Example 5 at $D_+ = 0.10$, and 18.6 times fewer on the long-range instance at $p = 0.05$ (Table 6; these are the parameters used in Section 5, and the advantage grows with $-\log \phi_{1/2}$ in both families). The informative-degree-max control tracks affinity-max where witnesses are a majority of the neighborhood, and is 10.4 times worse at the same long-range point, where they are not.

Refinement budget. Under unrestricted refinement these differences do not reach the output at $n = 10^4$, consistent with the robustness mechanism underlying Corollary 32 (Figure 2, left, thin curves). To see whether they nonetheless differ in robustness, we restrict the refinement sweep to the $\beta \log n$ label changes with the largest likelihood gain, ranked without reference to the ground truth. At $\beta = 0.3$ the rules separate: on Example 5 at $\lambda = 0.30$, where all three recover every instance unrestricted, affinity-max reaches 0.65 against 0.15 for nearest-first ($p = 0.006$), and on the long-range instance at $\lambda = 1.55$, 1.00 against 0.00 ($p < 10^{-5}$); inf.-degree-max again splits the two effects as above. Far enough above the transition Phase II makes no errors and every informed rule recovers under every budget tested. The full grid and the discordant counts are in Appendix B.2.

5.3 Why it works: suppressing the lower information tail

The right column of Figure 2 records $\mathcal{I}_S(u)$ at the moment each u is labeled, at $\lambda = 1.2\lambda_{\mathrm{IT}}$. The means agree across the three informed rules to within 0.03 nats, so no rule gathers more information on average; the lower tails differ by roughly a nat at the fifth percentile. Since the one-step bound of Lemma 25 is exponential in $\mathcal{I}_S(u)$, it is precisely this tail on which it depends: the adaptive rule does not accumulate more evidence, it avoids labeling vertices at which the evidence is weakest. A broader parameter sweep confirms the mechanism (Table 6).

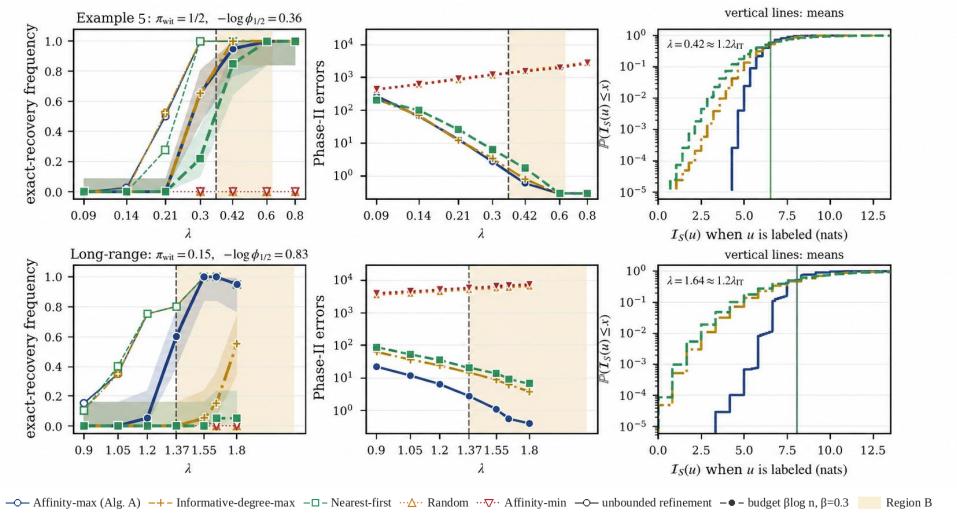


Figure 2: Phase-II selection rules on matched instances ($n = 10^4$, oracle initialization, 20 trials). Top: Example 5 ($\pi_{\text{wit}} = 1/2$, $-\log \phi_{1/2} = 0.36$); bottom: long-range instance ($\pi_{\text{wit}} = 0.15$, $-\log \phi_{1/2} = 0.83$). Left: exact recovery with unrestricted refinement (open) and budget $\beta = 0.3$ (filled). Middle: Phase-II errors. Right: CDF of $\mathcal{I}_S(u)$ at $1.2\lambda_{\text{IT}}$; vertical lines are means. Region B is shaded.

5.4 Practical initialization

Seed-corruption tolerance is consistent with the difference in π_{wit}: Example 5 survives 20% corruption, the long-range instance much less (Appendix B.3). Replacing exhaustive MAP by a localized spectral initializer on the seed region, inspired by (Allem et al., 2025), costs $O((\log n)^3)$ against an exponent of 25.2 for exhaustive search at the witness threshold of Example 5, and gives end-to-end recovery of 0.92 against 1.00 with the oracle on the long-range instance at $\lambda = 1.55$.

5.5 Comparison with block propagation

The block scale χ that the analysis of Algorithm B supplies is sized to make the auxiliary block structure percolate as $n \to \infty$: at $n = 10^4$ its blocks hold fewer than one vertex on average, and the algorithm does not propagate beyond the seed region over 220 instances. Enlarging χ beyond the range the analysis admits restores propagation at this size (Appendix B.4). We therefore compare against a *calibrated* arm, which selects the block scale empirically as implementations in this line of work do (Gaudio et al., 2024). We use the standard symmetric three-community GSBM ($P_{aa} = \text{Ber}(0.8)$, $P_{ab} = \text{Ber}(0.2)$, π uniform, $d = 2$, $r = 3$), the symmetric model of Remark 4 at $k = 3$, on which (Gaudio & Jin, 2026, Assumption 3) therefore fails identically while Assumption 1 holds comfortably ($\lambda_{\text{wit}} \approx 0.053$ against $\lambda_{\text{IT}} \approx 0.265$). Exact-recovery frequencies at $\lambda/\lambda_{\text{IT}} = 0.6, 0.7, 0.8, 0.9$ are $0.15, 0.30, 0.70, 1.00$ for Algorithm A against $0, 0.10, 0.55, 1.00$ for the calibrated baseline, a paired discordance of $11:1$ pooled over the three intermediate points (McNemar, $p = 0.006$). The two coincide by $0.9\lambda_{\text{IT}}$; the adaptive rule reaches the finite-sample transition slightly earlier without committing to a resolution. The grid lies below λ_{IT}, so this recovery is a finite-size effect. Calibration, constants and accuracy curves are in Appendix C.3.

6 Discussion and Open Problems

Algorithm A shows that spatial discretization is not operationally necessary for exact recovery in the GHCM: propagation can instead follow the available model-based information directly. Two questions remain open. Can the witness-connectivity condition, which our experiments suggest is a proof artifact for Algorithm A, be removed theoretically? And can block methods be designed with resolutions that remain effective at realistic finite sizes?

AI USE STATEMENT

We used generative AI tools to help design and implement the numerical experiments, including the construction of the long-range instance and the selection-rule controls, and to run and analyze the resulting simulations. We did not use generative AI tools to develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist with translation, or clean or reformat datasets.

Additionally, we used generative AI tools to generate initial draft ideas, text fragments and scientific figures, which were subsequently reviewed, verified, and rewritten by the authors. The mathematical claims and essential proof arguments were developed by the authors, and all AI-assisted mathematical exposition was checked against the authors' arguments. Synthetic data and figures produced with AI assistance were independently checked for consistency with the specified models and parameters. We take responsibility for the final content of this work, including all text, claims, data, figures, and other artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Complete proofs are included in the appendix. Anonymous code and deterministic random seeds for all numerical experiments in Section 5 and Appendices B–C are provided as supplementary material, together with archived trial-level outputs for the initialization and canonical GSBM experiments.

REFERENCES

Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms, 2015. URL https://arxiv.org/abs/1503.00609.

Emmanuel Abbe, Francois Baccelli, and Abishek Sankararaman. Community detection on Euclidean random graphs. *Information and Inference: A Journal of the IMA*, 10(1):109–160, 2021.

Luiz Emilio Allem, Konstantin Avrachenkov, Carlos Hoppen, Hariprasad Manjunath, and Lucas Siviero Sibemberg. Multi-community spectral clustering for geometric graphs. *arXiv preprint arXiv:2508.00893*, 2025.

Konstantin Avrachenkov, Andrei Bobu, and Maximilien Dreveton. Higher-order spectral clustering for geometric graphs. *Journal of Fourier Analysis and Applications*, 27(2):22, 2021.

Konstantin Avrachenkov, B. R. Vinay Kumar, and Lasse Leskelä. Community detection on block models with geometric kernels, 2024. URL https://arxiv.org/abs/2403.02802.

Sainyam Galhotra, Arya Mazumdar, Soumyabrata Pal, and Barna Saha. The geometric block model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Sainyam Galhotra, Arya Mazumdar, Soumyabrata Pal, and Barna Saha. Community recovery in the geometric block model. *Journal of Machine Learning Research*, 24:1–53, 2023.

Julia Gaudio and Andrew Jin. Exact recovery in the geometric SBM, 2025. URL https://arxiv.org/abs/2512.22773.

Julia Gaudio and Andrew Jin. Exact recovery in the geometric hidden community model, 2026. URL https://arxiv.org/abs/2601.17591.

Julia Gaudio, Xiaochun Niu, and Ermin Wei. Exact community recovery in the geometric SBM, 2024. URL https://arxiv.org/abs/2307.11196.

Julia Gaudio, Charlie Guan, Xiaochun Niu, and Ermin Wei. Exact label recovery in Euclidean random graphs, 2025. URL https://arxiv.org/abs/2407.11163.

G. Last and M. Penrose. *Lectures on the Poisson Process*. Cambridge University Press, 2018.

Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

Abishek Sankararaman and François Baccelli. Community detection on Euclidean random graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2181–2200. SIAM, 2018.

A NOTATION GUIDE

For convenience, Table 1 collects the notation used most frequently in the main text. Proof-specific constants are introduced locally.

Symbol	Meaning	Defined in
Model and information		
$\mathcal{S}_{d,n}$	d-dimensional torus of volume n	Def. 1
$P_{ab}(y)$	edge-weight law for labels a, b at rescaled distance y	Def. 1
$\bar{P}_{ab}(x)$	distance-rescaled edge-weight law	Def. 1
$\phi_t(P, Q)$	Hellinger/Chernoff affinity	(1)
$D_+(P_a\|P_b)$	Chernoff–Hellinger divergence between communities a, b	(3)
$W_{ab}(I)$	communities distinguishing a from b on interval I	Sec. 2
Thresholds and witness geometry		
c_d	connectivity constant (1 for $d \geq 2$, 2 for $d = 1$)	Assump. 1
λ_{IT}	information-theoretic exact-recovery threshold	Sec. 2
λ_{wit}	witness-connectivity threshold	Sec. 2
ℓ	common informative interval length	(7)
I_{ab}	informative interval chosen for pair a, b	(7)
π_{wit}	minimum witness mass over community pairs	Sec. 2
Algorithms		
B_{init}	seed region used for initialization	Sec. 3
$N(u)$	visible neighborhood of vertex u	Sec. 3
$\mathcal{I}_S(u)$	worst-case affinity information available to u from S	Def. 2
χ	spatial block scale used by Algorithm B	App. E
δ	minimum block-occupancy parameter	App. E

Table 1: Notation used repeatedly in the main text.

B EXPERIMENTAL DETAILS

B.1 MODELS, CONSTANTS AND THRESHOLDS

Example 5. $d = 2$, $r = 3$, $k = 3$, $\pi = (\frac{1}{2}, \frac{3}{10}, \frac{1}{5})$ and the kernel of (6). Witnesses are the two members of the pair, so $W_{ab}(I) = \{a, b\}$ on the informative bands and $\pi_{\mathrm{wit}} = \min_{a\neq b}(\pi_a + \pi_b) = \frac{1}{2}$ with $\ell = r/3 = 1$. At $D_+ = 0.10$ we have $p = 0.857$, $-\log \phi_{1/2} = 0.36$, $\lambda_{\mathrm{IT}} = 0.354$ and $\lambda_{\mathrm{wit}} = 0.637$. For $a \neq b$ and $c \notin \{a, b\}$ we have $P_{ac}(y) = P_{bc}(y) = \mathrm{Ber}(1-p)$ for *every* $y \in [0, r]$, so the third community is blind to the pair $\{a, b\}$ at all distances; the remaining two communities are witnesses exactly on the informative bands, since $\phi_{1/2}\big(\mathrm{Ber}(p), \mathrm{Ber}(1-p)\big) = 2\sqrt{p(1-p)} < 1$. Hence

$$W_{ab}(I) = \{a, b\} \quad \text{for } I \subseteq [0, r/3) \text{ or } I \subseteq (2r/3, r], \qquad W_{ab}(I) = \emptyset \text{ otherwise},$$

up to null sets. The supremum in the definition of λ_{wit} is attained at $\ell = r/3 = 1$, the length of a single informative band, since no interval of length $\ell > 1$ lies inside one band and $\ell^d \pi_{\mathrm{wit}}$ increases in ℓ. Assumption 1 therefore reads $\lambda \nu_d \ell^d \pi_{\mathrm{wit}} = \frac{1}{2}\pi\lambda > 1$, i.e. $\lambda_{\mathrm{wit}} = 2/\pi$. A direct computation with $t = \frac{1}{2}$, which is optimal by symmetry, gives

$$D_+(p) := \min_{a\neq b} D_+(P_a\|P_b) = \tfrac{2}{3}\Big(\min_{a\neq b}(\pi_a + \pi_b)\Big)\big(1 - 2\sqrt{p(1-p)}\big) = \tfrac{1}{3}\big(1 - 2\sqrt{p(1-p)}\big),$$

the factor $\frac{2}{3}$ being the mass that the distance density $2y/r^2$ places on the two informative bands. Condition (5) becomes $9\pi\lambda D_+(p) > 1$. Assumption 2 holds with $r_{\mathrm{init}} = r/3$ and $\Phi_{\mathrm{init}} = 2\sqrt{p(1-p)}$, and Assumption 3 with any $\rho > |\log(p/(1-p))|$. The two thresholds $\lambda > 2/\pi \approx 0.637$ and $9\pi\lambda D_+(p) > 1$ do not coincide, and since $D_+(p) \leq \frac{1}{3}$ the intermediate Region B of Figure 1 is non-empty for $\lambda \in \big(\frac{1}{3\pi}, \frac{2}{\pi}\big)$.

Long-range instance. $d = 2$, $r = 3$, $k = 3$, $\pi = (0.50, 0.35, 0.15)$. Observations are informative only on the shell $[2.5, 3]$, where $P_{13}(y) = \mathrm{Ber}(p)$, $P_{23}(y) = \mathrm{Ber}(1-p)$ and every other pair is $\mathrm{Ber}(1/2)$; a weak band on $[0, 0.25)$ with contrast $\varepsilon = 0.02$ makes Assumption 2 hold with $\Phi_{\mathrm{init}} = \sqrt{1 - 4\varepsilon^2}$. Table 2 gives the kernel in full.

Table 2: The long-range instance: $P_{ab}(y)$ for all $a, b \in \{1, 2, 3\}$ and all $y \in [0, r]$, with $r = 3$, $\varepsilon = 0.02$ and $p = 0.05$. Entries are Bernoulli parameters. Only the outer shell separates communities 1 and 2, and only through community 3.

pair $\{a, b\}$	weak band $y \in [0, 0.25)$	uninformative $y \in [0.25, 2.5)$	outer shell $y \in [2.5, 3]$
$\{1, 1\}$	$\frac{1}{2} + \varepsilon$	$\frac{1}{2}$	$\frac{1}{2}$
$\{2, 2\}$	$\frac{1}{2} + \varepsilon$	$\frac{1}{2}$	$\frac{1}{2}$
$\{3, 3\}$	$\frac{1}{2} + \varepsilon$	$\frac{1}{2}$	$\frac{1}{2}$
$\{1, 2\}$	$\frac{1}{2} - \varepsilon$	$\frac{1}{2}$	$\frac{1}{2}$
$\{1, 3\}$	$\frac{1}{2} - \varepsilon$	$\frac{1}{2}$	p
$\{2, 3\}$	$\frac{1}{2} - \varepsilon$	$\frac{1}{2}$	$1 - p$

Hence $W_{12}(I) = \{3\}$ for $I \subseteq [2.5, 3]$, while $W_{13}(I) = W_{23}(I) = Z$ there, so $\pi_{\mathrm{wit}} = \pi_3 = 0.15$ on the shell; on the weak band every pair has witness mass $\pi_a + \pi_b \geq 0.5$. The minimum in (5) is attained at $\{1, 2\}$: at $t = \frac{1}{2}$ the shell contributes $\pi_3\big(1 - 2\sqrt{p(1-p)}\big) = 0.0846$ against 0.1527 for the other two pairs, and the weak band contributes $(\pi_1 + \pi_2)\big(1 - \sqrt{1 - 4\varepsilon^2}\big)$. Weighting by the mass $2y/r^2$ places on each band, 0.00694 and 0.30556, gives $D_+ = 0.0259$. Hence $-\log \phi_{1/2} = 0.83$, $\lambda_{\mathrm{IT}} = 1.368$, and $\lambda_{\mathrm{wit}} = 8.49$ from the shell values above. Since $\lambda_{\mathrm{wit}} \approx 8.49$ exceeds every simulated λ, Proposition 7 gives that no admissible block configuration exists on this instance at any point of the grid: every simulated point above λ_{IT} is one at which the block comparator has no certified resolution, while Algorithm A runs unchanged.

The supremum in Assumption 1 is worth stating explicitly for this instance, since it is not attained where witness mass is largest. On the weak band every pair has witness mass at least 0.5, but the band has length 0.25; on the shell the pair $\{1, 2\}$ has witness mass only 0.15, but the band has length 0.5. Maximizing $\ell^d \pi_{\mathrm{wit}}$ gives $0.25^2 \cdot 0.5 = 0.031$ against $0.5^2 \cdot 0.15 = 0.0375$, so the supremum is attained on the shell with $\ell = 1/2$ and $\pi_{\mathrm{wit}} = 0.15$.

B.2 Ordering ablation in full

Phase-II error counts across the full λ grid for both instances are the middle column of Figure 2, reported as the total over the 20 matched instances at each grid point. Table 3 gives the McNemar discordant counts underlying the budgeted comparisons of Section 5.2. The witness-informativeness sweep referenced in Section 5.3 is Table 6, in Appendix C.2.

Table 3: McNemar exact test on discordant instances, at the grid points and budget reported in Section 5.2 ($n = 10\,000$, 20 trials, matched instances). The entry $a\!:\!b$ gives the number of instances the first rule recovers and the second does not, against the reverse.

instance	λ	comparison	discordant	p
Example 5	0.30	affinity-max vs. nearest-first	11 : 1	0.006
	0.30	affinity-max vs. inf.-degree-max	4 : 4	1.0
long-range	1.55	affinity-max vs. nearest-first	20 : 0	$< 10^{-5}$
	1.55	inf.-degree-max vs. nearest-first	19 : 0	$< 10^{-5}$

Two features are worth noting. First, the discordance is almost perfectly one-sided: across all grid points and budgets, affinity-max loses at most one instance to nearest-first where nearest-first wins many. Second, the separation is monotone in the budget at fixed λ, and vanishes for every budget once

λ is far enough above the transition that Phase II makes no errors at all — $\lambda \geq 0.42$ on Example 5 and $\lambda \geq 1.8$ on the long-range instance.

B.3 INITIALIZATION

Seed tolerance. Table 4 reports end-to-end exact recovery when a fraction ε of the seed labels is corrupted. The long-range instance, with smaller π_{wit}, tolerates less corruption.

Table 4: End-to-end exact-recovery frequency when a fraction ε of the seed labels is corrupted to uniformly random incorrect labels ($n = 10\,000$; 25 trials for Example 5, 20 for the long-range instance). Tolerance tracks the witness mass π_{wit}.

instance	λ	$\varepsilon = 0$	0.02	0.05	0.10	0.20	0.30
Example 5 ($\pi_{\mathrm{wit}} = \frac{1}{2}$)	0.30	1.00	0.96	1.00	0.96	0.92	0.48
	0.42	1.00	1.00	1.00	1.00	1.00	0.88
long-range ($\pi_{\mathrm{wit}} = 0.15$)	1.37	0.85	0.85	0.80	0.75	0.55	0.10
	1.55	1.00	1.00	0.95	0.95	0.75	0.40

The initializer. Let $S = V \cap B_{\mathrm{init}}$ with $m = |S|$ and let X be the $m \times m$ matrix of observed weights on seed-internal visible pairs. Put $p_{ab}(y) = \mathbb{E}[X_{uv}]$ under $P_{ab}(y)$ and let

$$q_0(y) \;=\; \sum_{a,b \in Z} \pi_a \pi_b p_{ab}(y),$$

and set $C_{uv} = X_{uv} - q_0(\|u - v\|/(\log n)^{1/d})$ on visible pairs and $C_{uv} = 0$ otherwise. Since $\mathbb{E}[C \mid \sigma]_{uv} = p_{\sigma_u \sigma_v}(y) - q_0(y)$ with $y = \|v - u\|/(\log n)^{1/d}$ is a general $k \times k$ structure, the construction does not require the model to be of the same-community / different-community form, and we apply it to both instances. We take the k eigenvectors of C with largest $|\lambda|$, cluster their rows by k-means with 12 restarts, discard restarts that produce an empty cluster, score the survivors by the GHCM seed likelihood after resolving the $k!$ cluster-to-label assignments, and apply the refinement of Algorithm 3 to the seed. Scoring restarts by the seed likelihood rather than by k-means inertia matters: the failure mode is a collapsed cluster, which inertia accepts and the likelihood does not. The outcome is Table 5.

Table 5: Localized spectral initializer in place of the oracle, $n = 10\,000$, 25 trials. Seed accuracy understates usefulness: on the long-range instance the median seed error is 1 vertex in 510, and Phase II–III absorb it.

instance	λ	seed exact	median seed errors	end-to-end exact	oracle end-to-end
Example 5	0.30	0.56	0/100	0.76	0.96
	0.42	0.76	0/139	0.76	1.00
long-range	1.37	0.28	3/451	0.72	0.92
	1.55	0.32	1/510	0.92	1.00

Finite-size behavior. Exact seed recovery improves with n while the seed itself grows only logarithmically. On Example 5 at $\lambda = 0.42$ it rises from 0.60 at $n = 10^3$ ($m \approx 101$) to 0.88 at $n = 10^5$ ($m \approx 177$); at $\lambda = 0.637$ from 0.80 ($m \approx 155$) to 0.96 ($m \approx 267$). Failures are bimodal: either a handful of vertices, which Phase II–III absorb, or a collapsed cluster affecting 40–100% of the seed, which exceeds the tolerance of Table 4. The collapse rate falls with n, from 6 instances in 25 at $n = 10^3$ to 3 in 25 at $n = 10^5$ on Example 5 at $\lambda = 0.42$, and accounts for the entire residual gap in Table 5.

Tie-breaking. The kernels considered here produce exact ties in the local log-likelihood: a vertex whose labeled evidence is symmetric in two communities has two identical scores. We break such ties deterministically, to the lowest index, so that the reported labels do not depend on the order in which the likelihood sum is accumulated.

B.4 BLOCK COMPARATOR

Certification (14) requires sufficiently many labeled vertices from the witness set of every community pair; for Example 5 the bottleneck is the combined mass of the two rarest communities, while blocks hold $\lambda\ell^d\chi\log n$ vertices in expectation. The admissible χ is constrained from above by (9) and (10): the second forces $\sqrt{d+3}\,\chi^{1/d} < 1$, and the first binds well before that, since its left-hand side decreases in χ and already approaches $\lambda\pi_{\mathrm{wit}}\nu_d\ell^d$ as $\chi \downarrow 0$. Near the witness threshold, where $\lambda\nu_d\ell^d\pi_{\mathrm{wit}}$ only just exceeds c_d, the admissible range of χ is therefore small, and the expected occupancy $\lambda\ell^d\chi\log n$ is correspondingly small.

Since occupancy grows only logarithmically in n, the obstruction persists to extremely large problem sizes: raising n by several orders of magnitude changes the expected number of vertices per block by a constant factor. The constants are asymptotic: χ and δ are sized so that Proposition 24 holds as $n \to \infty$, and nothing in the analysis claims they are the right resolution at any finite n. On Example 5 at $\lambda = 0.8$, the only simulated point above $\lambda_{\mathrm{wit}} \approx 0.637$, (9) forces $\chi < 2.6 \times 10^{-3}$, so the expected occupancy $\lambda\ell^d\chi\log n$ is below 0.02 vertices per block at $n = 10^4$ against a certification threshold of $\frac{1}{2}\delta\pi_{\mathrm{wit}}\log n$. This is why Algorithm B does not propagate beyond the seed region over 220 instances. Tuning χ empirically restores propagation at finite n, which is what the calibrated arm of Appendix C.3 does.

C ADDITIONAL EXPERIMENTAL RESULTS

C.1 PHASE DIAGRAM FOR ALGORITHM A ON EXAMPLE 5

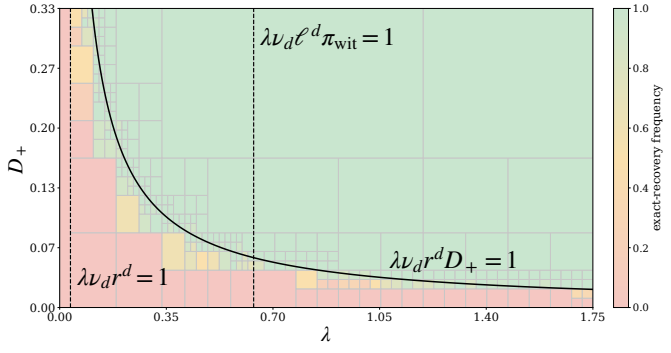


Figure 3: Empirical exact-recovery frequency of Algorithm A followed by refinement, $n = 10\,000$, oracle initialization. Blocks meeting the exact-recovery boundary are subdivided recursively up to five levels; 20 trials along the boundary and 10 elsewhere. The solid curve is $\lambda\nu_d r^d D_+ = 1$; the left dashed line is $\lambda\nu_d r^d = 1$ and the right dashed line the witness-connectivity threshold $\lambda\nu_d\ell^d\pi_{\mathrm{wit}} = 1$. The empirical boundary tracks the information-theoretic curve.

C.2 WITNESS-INFORMATIVENESS SWEEP

Table 6: Phase-II errors summed over 20 matched instances, each row evaluated at $\lambda = 1.2\lambda_{\mathrm{IT}}$. Here $\mu = \log n/(1-\phi_{1/2})$ is the expected number of witnesses per vertex at threshold. The advantage over nearest-first grows with $-\log\phi_{1/2}$, while the advantage over inf.-deg.-max is governed by π_{wit}.

				Phase-II errors			ratio	
model	param.	$-\log\phi_{1/2}$	μ	aff.	inf.-deg.	near.	near./aff.	inf.-deg./aff.
Example 5	$D_+ = 0.05$	0.16	61.4	35	46	61	1.7	1.3
($\pi_{\mathrm{wit}} = \frac{1}{2}$)	$D_+ = 0.10$	0.36	30.7	15	12	29	1.9	0.8
	$D_+ = 0.20$	0.92	15.4	4	2	20	5.0	0.5
	$D_+ = 0.30$	2.30	10.2	2	3	54	27.0	1.5
long-range	$p = 0.10$	0.51	23.0	57	219	294	5.2	3.8
($\pi_{\mathrm{wit}} = 0.15$)	$p = 0.05$	0.83	16.3	11	114	205	18.6	10.4
	$p = 0.02$	1.27	12.8	7	114	190	27.1	16.3

C.3 CANONICAL GSBM BENCHMARK

Model and constants. $d = 2$, $r = 3$, $k = 3$, $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $P_{aa} = \mathrm{Ber}(0.8)$ and $P_{ab} = \mathrm{Ber}(0.2)$ for $a \neq b$, independent of distance on $[0, r]$. Then $\phi_{1/2}(\mathrm{Ber}(0.8), \mathrm{Ber}(0.2)) = 0.8$ and, since the third community is pair-blind,

$$D_+ = 1 - \tfrac{1}{3}\big(0.8 + 0.8 + 1\big) = \tfrac{2}{15}, \qquad \lambda_{\mathrm{IT}} = \frac{1}{\nu_d r^d D_+} = \frac{5}{6\pi} \approx 0.265.$$

Witnesses are the two members of the pair, so $\pi_{\mathrm{wit}} = \frac{2}{3}$ and $\ell = r$ is admissible, giving $\lambda_{\mathrm{wit}} = 1/(6\pi) \approx 0.053$.

Calibration. The block baseline has one free parameter, the target expected occupancy, which we select on calibration instances disjoint from the test instances. For each λ we run 6 calibration instances over the occupancy grid $\{1.5, 2, 2.5, 3, 4, 5, 8, 10\}$ and retain the value with the highest exact-recovery count, breaking ties towards the smaller value. The selected occupancies are 1.5, 2.0 and 2.5 at $\lambda/\lambda_{\mathrm{IT}} = 0.7, 0.8, 0.9$; at the two lowest grid points no occupancy recovers any calibration instance, and we carry the value selected at 0.7. That the selection moves with λ is the point: the baseline needs a resolution, and the right resolution is not a constant of the model.

Results. Table 7 reports exact-recovery frequency and mean accuracy on the test instances. Mean accuracy separates the two arms well below the transition, where neither recovers exactly.

Table 7: Symmetric three-community GSBM at $n = 10^4$, matched instances (20 trials per point, 10 at $\lambda/\lambda_{\mathrm{IT}} = 0.5$ and 0.9). Every grid point lies below $\lambda_{\mathrm{IT}} = 0.265$, so recovery here is a finite-size effect.

$\lambda/\lambda_{\mathrm{IT}}$	0.5	0.6	0.7	0.8	0.9
λ	0.133	0.159	0.186	0.212	0.239
exact-recovery frequency					
Algorithm A	0.00	0.15	0.30	0.70	1.00
calibrated block	0.00	0.00	0.10	0.55	1.00
mean accuracy					
Algorithm A	0.954	0.997	0.999	1.000	1.000
calibrated block	0.905	0.933	0.980	1.000	1.000

Paired discordant counts (Algorithm A recovers, block does not, against the reverse) are 3 : 0, 5 : 1 and 3 : 0 at $\lambda/\lambda_{\mathrm{IT}} = 0.6, 0.7, 0.8$, and 0 : 0 at both endpoints. McNemar's exact test gives $p = 0.25$, 0.22 and 0.25 individually and $p = 0.006$ pooled over the three; we report the pooled figure and note that no single grid point is significant on its own.

On direct spectral recovery. We also applied the multi-community spectral method of Allem et al. (2025) directly to these instances, rather than to the seed region as in Section 5.4. It did not cluster at this sparsity: mean accuracy stayed between 0.374 and 0.399 across the grid, against $1/3$ for a uniformly random labeling. We do not report this as a comparison, since a baseline at chance level measures our implementation instead of the method, whose analysis targets a denser regime; we note only that we did not find a direct application that worked here, and that the same construction is effective as a seed initializer, where it operates on $O((\log n)^2)$ observations within B_{init}.

D Initialization, Propagation and Refinement algorithms

Algorithm 1 Maximum a Posteriori

Input: Graph G and vertex set $S \subseteq V$.
Output: An estimated labeling $\widehat{\sigma}_S : S \to Z$.

1: Set

$$\widehat{\sigma}_S \in \operatorname*{arg\,max}_{\sigma: S \to Z} \left\{ \sum_{u \in S} \log \pi_{\sigma(u)} + \sum_{\{u,v\} \subseteq S} \log \bar{p}_{\sigma(u)\sigma(v)}(x_{uv}; \|u - v\|) \right\}.$$

Algorithm 2 Propagate

Input: Graph G, S, S' mutually disjoint vertex sets and a labeling $\widehat{\sigma}_S : S \to Z$.
Output: An estimated labeling $\widehat{\sigma}_{S'} : S' \to Z$.

1: For each $u \in S'$ set

$$\widehat{\sigma}_{S'}(u) = \operatorname*{arg\,max}_{b \in Z} \sum_{v \in S : u \leftrightarrow v} \log \bar{p}_{\widehat{\sigma}_S(v) b}(x_{uv}; \|u - v\|).$$

Algorithm 3 Refine

Input: Graph G, a vertex $u \in V$, and a labeling $\widehat{\sigma} : V \to Z \sqcup \{*\}$.
Output: An estimated label $\widetilde{\sigma}(u) \in Z$.

1: Set

$$\widetilde{\sigma}(u) = \operatorname*{arg\,max}_{a \in Z} \sum_{v \in V \setminus \{u\} : u \leftrightarrow v, \widehat{\sigma}(v) \neq *} \log \bar{p}_{a \widehat{\sigma}(v)}(x_{uv}; \|u - v\|).$$

E Block-Based Comparator

The adaptive algorithm above is our main method. For comparison, we also analyze a block-based Algorithm B, generalizing Gaudio & Jin (2026) (and the propagation scheme originating in Gaudio et al. (2024)). This second algorithm keeps the spatial discretization while incorporating the same pair-specific witness idea. Both algorithms are analyzed under the same witness-connectivity condition; the distinction is that the block algorithm uses the partition operationally, whereas for the adaptive method the partition appears only in the proof. Recall ℓ, the intervals I_{ab}, and π_{wit} from Section 2, and in particular that $\lambda \nu_d \ell^d \pi_{\mathrm{wit}} > c_d$ by (7). Next we describe how we partition $\mathcal{S}_{d,n}$. If $d = 1$, take

$$0 < \chi < \frac{1}{2}\Big(1 - \frac{1}{\lambda \pi_{\mathrm{wit}} \ell}\Big). \tag{8}$$

If $d \geq 2$, take $\chi > 0$ such that

$$\lambda \pi_{\mathrm{wit}} \ell^d \Big(\nu_d \Big(1 - \frac{3\sqrt{d}}{2} \chi^{1/d}\Big)^d - \chi \Big) > 1 \quad \text{and} \quad 1 - \frac{3\sqrt{d}}{2} \chi^{1/d} > 0. \tag{9}$$

The existence of χ is guaranteed by Lemma 19. By decreasing χ further if necessary, we also assume throughout that

$$\ell \chi^{1/d} < \frac{r}{2} \qquad \text{and} \qquad \sqrt{d+3}\, \chi^{1/d} < 1. \tag{10}$$

Since the preceding conditions on χ are strict and remain valid for all sufficiently small positive χ, this entails no additional assumption on the model parameters. We discretize $\mathcal{S}_{d,n}$ into a family $\mathcal{P}_\chi$ of cubes of volume $\ell^d \chi \log n$, which we call *blocks*. For $s > 0$ we call two blocks B_i and B_j *s-close* if

$$\sup_{u \in B_i, v \in B_j} \|u - v\| \leq s(\log n)^{1/d}, \tag{11}$$

where $\|\cdot\|$ is the toroidal metric on $\mathcal{S}_{d,n}$. For $B, B' \in \mathcal{P}_\chi$ we write $B \sim B'$ if B and B' are ℓ-close with ℓ given at (7), and $B \leftrightarrow B'$ if they are r-close. For $d \geq 2$ let K_d denote the number of blocks that are ℓ-close to a given block B, excluding the block itself. If $d = 1$, let K_d denote the number of blocks that are ℓ-close to the left (or to the right) to a given block, excluding the block itself. Given a graph G on $\mathcal{S}_{d,n}$ and a partition of $\mathcal{S}_{d,n}$ into blocks of volume $v(n)$, we follow (Gaudio & Jin, 2026, Definition 5) and define the $(v(n), c(n))$-block visibility graph of G as the graph $H = (V^\dagger, E^\dagger)$, where $V^\dagger := \{i \in [n/v(n)] : V(B_i) \geq c(n)\}$ is the set of all blocks with at least $c(n)$ vertices and $E^\dagger := \{\{i, j\} : i, j \in V^\dagger, B_i \sim B_j\}$ is the set of all pairs of blocks in $V^\dagger$ that are mutually ℓ-close. Let $\rho > 0$ satisfy Assumption 3, define

$$\begin{aligned} \delta_0 &:= \frac{\left(\lambda \pi_{\mathrm{wit}} \ell^d \chi K_d - 1\right)^2}{8 \lambda \pi_{\mathrm{wit}} \ell^d \chi K_d^2}, \\ \beta_0 &:= \frac{\lambda \nu_d r^d \min_{a \neq b} D_+(P_a \| P_b) - 1}{4\rho}, \\ K_r &:= \left\lceil \frac{\nu_d \left(r + \sqrt{d}\, \ell \chi^{1/d}\right)^d}{\ell^d \chi} \right\rceil. \end{aligned} \tag{12}$$

and fix

$$0 < \delta < \min\left\{\delta_0, \frac{\beta_0}{K_r}\right\}. \tag{13}$$

We call a block B δ-occupied if $V(B) \geq \delta \log n$. The parameters χ and δ are chosen so that the blocks contain enough nodes of each community and so that the $(\ell^d \chi \log n, \delta \log n)$-block visibility graph $H = (V^\dagger, E^\dagger)$ of δ-occupied blocks is connected with high probability.

Given a labeling $\hat{\sigma}$, we call a δ-occupied block B *$\hat{\sigma}$-certified* if

$$\min_{a \neq b} \sum_{w \in W_{ab}(I_{ab})} |\{v \in V \cap B : \hat{\sigma}(v) = w\}| \geq \frac{1}{2} \delta \pi_{\mathrm{wit}} \log n. \tag{14}$$

If δ and $\hat{\sigma}$ are clear from the context, let $\hat{V}^\dagger_{\mathrm{cert}} \subseteq V^\dagger$ denote the subset of $\hat{\sigma}$-certified blocks.

We next give an overview of our second, block-based algorithm, which generalizes the algorithm given in Gaudio & Jin (2026). We start performing a Maximum A Posteriori estimate on the initial set B_{init}, which is identical to Algorithm A. Then we divide $\mathcal{S}_{d,n}$ into blocks of volume $\ell^d \chi \log n$ and construct the $(\ell^d \chi \log n, \delta \log n)$-visibility graph $H = (V^\dagger, E^\dagger)$ of δ-occupied blocks. Given that H is connected, we determine the subset of all $\hat{\sigma}$-certified blocks. We then choose a δ-occupied block B which is not fully contained in B_{init} and ℓ-close to a $\hat{\sigma}$-certified block. We apply Propagate (Algorithm 2) on input $(G, \bigcup_{k \in \hat{V}^\dagger_{\mathrm{cert}}} V \cap B_k, V \cap B_i, \hat{\sigma})$ to obtain a labeling $\hat{\sigma}$ on the vertex set of B. We then update $\hat{V}^\dagger$ and $\hat{V}^\dagger_{\mathrm{cert}}$ and apply Propagate on the next δ-occupied block which is ℓ-close to a $\hat{\sigma}$-certified block. We will show that this procedure allows us to label all blocks in H with high probability. Such a technique of *propagation* was first introduced in 2024 by Gaudio et al. (2024). The difference to the algorithm in Gaudio & Jin (2026) is that in the propagation step we do not only use the estimated labeling in one parent block, but from all certified blocks intersecting the visibility range of a given block. This is necessary, since we work under the pair-specific witness Assumption 1, which only guarantees that a given pair of communities can be distinguished in a small interval.

Algorithm B Block-based Exact Recovery

Input: $G \sim \mathrm{GHCM}(\lambda, n, r, \pi, P(\cdot), d)$.
Output: An estimated community labeling $\widetilde{\sigma} : V \to Z$.

Phase I:
Let $\hat{V} := V \cap B_{\mathrm{init}}$.
Apply `Maximum a Posteriori` (Algorithm 1) on input $(G, \hat{V})$ to obtain a labeling $\hat{\sigma}$ on $\hat{V}$.
Phase II:
Take χ satisfying (8) if $d = 1$, or (9) if $d \geq 2$, and $\delta > 0$ satisfying (13).
Partition $\mathcal{S}_{d,n}$ into $n/(\ell^d \chi \log n)$ blocks of volume $\ell^d \chi \log n$.
Let B_i be the i-th block for $i \in [\lceil n/(\ell^d \chi \log n) \rceil]$.
Construct the $(\ell^d \chi \log n, \delta \log n)$-visibility graph $H = (V^\dagger, E^\dagger)$ of δ-occupied blocks.
Let $\hat{V}^\dagger := \{i \in V^\dagger : B_i \subseteq B_{\mathrm{init}}\}$ and let $\hat{V}^\dagger_{\mathrm{cert}} \subseteq \hat{V}^\dagger$ be the subset of all $\hat{\sigma}$-certified blocks.
while $\{k \in V^\dagger \setminus \hat{V}^\dagger : \exists j \in \hat{V}^\dagger_{\mathrm{cert}} : B_j \sim B_k\} \neq \emptyset$ **do**
    Let $i := \min\{k \in V^\dagger \setminus \hat{V}^\dagger : \exists j \in \hat{V}^\dagger_{\mathrm{cert}} : B_j \sim B_k\}$.
    Apply `Propagate` on input $(G, \bigcup_{k \in \hat{V}^\dagger_{\mathrm{cert}}} (V \cap B_k), (V \cap B_i) \setminus \hat{V}, \hat{\sigma})$ to obtain labels on $(V \cap B_i) \setminus \hat{V}$.
    $\hat{V} \leftarrow \hat{V} \cup (V \cap B_i)$.
    $\hat{V}^\dagger \leftarrow \hat{V}^\dagger \cup \{i\}$.
    if B_i is $\hat{\sigma}$-certified **then** $\hat{V}^\dagger_{\mathrm{cert}} \leftarrow \hat{V}^\dagger_{\mathrm{cert}} \cup \{i\}$.
for $v \in V \setminus \hat{V}$ **do**
    Set $\hat{\sigma}(v) = *$.
Phase III:
for $v \in V$ **do**
    Apply `Refine` (Algorithm 3) on input $(G, v, \hat{\sigma})$ to compute $\tilde{\sigma}(v)$.

F Theoretical background

F.1 Model conventions

Edge weights take values in a measurable space $(\mathbb{X}, \mathcal{X})$ carrying a σ-finite measure μ. We assume $P_{ij}(y) \ll \mu$ for all $y \in \mathbb{R}$ and write $p_{ij}(\cdot\,; y)$ for the density of $P_{ij}(y)$ with respect to μ, and $\bar{p}(\cdot\,; y)$ for that of $\bar{P}(y)$. The visibility radius of Definition 1 is

$$r := \inf\{y > 0 : \exists x \in \mathbb{X} \text{ such that } P_{ab}(y') = \delta_x \text{ for all } y' > y \text{ and } a, b \in Z\} < \infty,$$

with δ_x the Dirac measure at x: beyond $r(\log n)^{1/d}$ every pair produces the same deterministic weight, so such pairs carry no information about the labels. For $B \subseteq \mathcal{S}_{d,n}$ measurable we write $V(B)$ for the number of vertices in B, and $V_i := \{v \in V : \sigma^*(v) = i\}$, $V_i(B) := |V_i \cap B|$.

F.2 Preliminaries

The following Mecke equation is a standard tool in the analysis of Poisson processes. For a proof see (Last & Penrose, 2018, Theorem 4.1).

Proposition 8 (Mecke equation)**.** *For any measurable $f \geq 0$,*

$$\mathbb{E}\Big[\sum_{v \in V} f(v, V)\Big] = \lambda \int_{\mathcal{S}_{d,n}} \mathbb{E}[f(v, V \cup \{v\})]\, \mathrm{d}v. \tag{15}$$

The following lemma provides standard concentration inequalities. The statement is taken from Mitzenmacher & Upfal (2005) and its proof can be found therein.

Lemma 9 (Poisson Chernoff bound)**.** *Let $X \sim \mathrm{Poisson}(\lambda)$ for some $\lambda > 0$. Then we have*

$$\mathbb{P}(X \geq x) \leq \frac{(e\lambda)^x e^{-\lambda}}{x^x}, \qquad x > \lambda,$$
$$\mathbb{P}(X \leq x) \leq \frac{(e\lambda)^x e^{-\lambda}}{x^x}, \qquad x < \lambda.$$

Lemma 9 allows us to find an upper bound for the number of nodes in a block.

Corollary 10. *There exists a constant $\Delta > 0$ such that, with high probability, for all blocks B_i, we have $V(B_i) \leq \Delta \log n$.*

Proof. Note that $V(B_i)$ is Poisson($\lambda\chi\ell^d \log n$)-distributed. By Lemma 9 and a union bound, we have for all $\Delta > \lambda\chi\ell^d$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\lceil n/(\ell^d \chi \log n)\rceil} \{V(B_i) > \Delta \log n\}\right) \leq O\Big(\frac{n}{\log n}\Big) n^{-\lambda\chi\ell^d} \Big(\frac{e\lambda\chi\ell^d}{\Delta}\Big)^{\Delta \log n}.$$

Since $\Delta \log\left(\frac{e\lambda\chi\ell^d}{\Delta}\right) \to -\infty$ as $\Delta \to \infty$, the assertion follows. □

The following lemma will be useful in the proof of exact recovery of the initialization step and almost-exact recovery of the propagation step. It generalizes (Gaudio & Jin, 2026, Lemma 5).

Lemma 11 (Informative interval)**.** *Let $a, b, a', b' \in Z$ and $t \in (0, 1)$. Suppose that there exists an interval $I \subseteq [0, r]$ of positive length such that*

$$P_{ab}(y) \neq P_{a'b'}(y) \qquad \text{for a.e. } y \in I.$$

Then, for every $\varepsilon > 0$, there exists $\gamma \in (0, 1)$ such that

$$\big|\big\{y \in I : \phi_t\big(P_{ab}(y), P_{a'b'}(y)\big) \geq 1 - \gamma\big\}\big| < \varepsilon |I|.$$

Proof. We skip the proof since it is almost identical to the proof of (Gaudio & Jin, 2026, Lemma 5). □

Let $B_r := \{x \in \mathbb{R}^d : \|x\| \leq r\}$. For any interval $I \subset [0, r]$ define the annulus

$$\mathcal{A}_I := \Big\{u \in B_r : \|u\| \in I\Big\}. \tag{16}$$

The following corollary is a direct consequence of Lemma 11. It is the analogue of (Gaudio & Jin, 2026, Corollary 1).

Corollary 12. *Let $a, b, a', b' \in Z$ and let $I \subseteq [0, r]$ be an interval of positive length such that*

$$P_{ab}(y) \neq P_{a'b'}(y) \qquad \text{for a.e. } y \in I.$$

Let D denote the distance from the origin of a point chosen uniformly at random from $\mathcal{A}_I$. Then, for every $\epsilon > 0$ and $t \in (0, 1)$, there exists $\gamma \in (0, 1)$ such that

$$\mathbb{P}\left(\phi_t\big(P_{ab}(D), P_{a'b'}(D)\big) \geq 1 - \gamma\right) \leq \epsilon.$$

The value of γ depends only on ϵ, t, and I.

G Exact recovery on the initial set

In this section we establish exact recovery on the initial set $V_{\mathrm{init}} := V \cap B_{\mathrm{init}}$, where

$$B_{\mathrm{init}} = [-r(\log n)^{1/d}, r(\log n)^{1/d}]^d.$$

We follow the strategy of Gaudio et al. (2025); Gaudio & Jin (2026) and first analyze a restricted Maximum Likelihood Estimator (MLE). Specifically, let

$$\bar{\sigma} = \operatorname*{arg\,max}_{\sigma \in X_0^*(\epsilon)} \ell_0(G, \sigma),$$

where

$$\ell_0(G,\sigma) := \sum_{u\in V_{\text{init}}} \sum_{\substack{v\in V_{\text{init}}\\ v\neq u}} \log\left(\bar{p}_{\sigma(u),\sigma(v)}(x_{uv};\|u-v\|)\right)$$

and

$$X_0^*(\epsilon) := \left\{\sigma : V_{\text{init}} \to Z : |\{u\in V_{\text{init}} : \sigma(u)=j\}| \in \left((\pi_j-\epsilon)|V_{\text{init}}|, (\pi_j+\epsilon)|V_{\text{init}}|\right)\, \forall j\in Z\right\}. \tag{17}$$

We first show that $\bar{\sigma}$ achieves exact recovery on V_{init} with high probability.

Proposition 13. *Let $G\sim \text{GHCM}(\lambda,n,r,\pi,P(\cdot),d)$ and let $\bar{\sigma}$ be the restricted MLE. Suppose that Assumptions 1, 2 and 3 hold. Then, $\bar{\sigma}$ achieves exact recovery on V_{init} with high probability.*

Our proof of Proposition 13 is an adaptation of the proof of Lemma D.4 in Gaudio et al. (2025).

Lemma 14. *For any $0<\epsilon<\min_{j\in Z}\pi_j/2$, it holds that $\mathbb{P}(\sigma^*\in X_0^*(\epsilon)) = 1-o(1)$.*

Proof. The number of vertices in the initial cube with community label j is an independent Poisson random variable $N_j \sim \text{Poisson}(\lambda\pi_j(2r)^d\log n)$, and $|V_{\text{init}}| = \sum_{j\in Z} N_j \sim \text{Poisson}(\mu)$ where $\mu := \lambda(2r)^d\log n$. We show that $N_j \in ((\pi_j-\epsilon)|V_{\text{init}}|, (\pi_j+\epsilon)|V_{\text{init}}|)$ for all $j\in Z$ with high probability.

First we show that $|V_{\text{init}}|$ concentrates around μ. By Lemma 9 and the elementary inequality $(e/(1+\epsilon))^{1+\epsilon} \le e^{1-\epsilon^2/4}$ for all $\epsilon\in(0,1)$,

$$\begin{aligned}\mathbb{P}\Big(||V_{\text{init}}|-\mu| \ge \frac{\epsilon}{4}\mu\Big) &\le \mathbb{P}\Big(|V_{\text{init}}| \ge (1+\frac{\epsilon}{4})\mu\Big) + \mathbb{P}\Big(|V_{\text{init}}| \le (1-\frac{\epsilon}{4})\mu\Big)\\ &\le 2e^{-\mu}\Big(\frac{e}{1+\frac{\epsilon}{4}}\Big)^{(1+\epsilon/4)\mu} \le 2e^{-\mu\epsilon^2/64} = o(1).\end{aligned}$$

Second, we show that N_j concentrates around $\pi_j\mu$ for each $j\in Z$,

$$\mathbb{P}\Big(|N_j-\pi_j\mu| \ge \frac{\epsilon}{4}\pi_j\mu\Big) \le 2\exp\left(-\frac{\epsilon^2\pi_j\mu}{64}\right) = o(1).$$

A union bound over all $j\in Z$ gives $N_j \in ((1-\epsilon/4)\pi_j\mu,\ (1+\epsilon/4)\pi_j\mu)$ for all j, with high probability.

Using the intersection of the two events above, we can upper bound the difference with high probability. For each $j\in Z$

$$\frac{N_j}{|V_{\text{init}}|} \le \frac{(1+\epsilon/4)\pi_j\mu}{(1-\epsilon/4)\mu} = \pi_j\cdot\frac{1+\epsilon/4}{1-\epsilon/4} \le \pi_j+\epsilon,$$

where the last inequality holds since $\epsilon<\min_{j\in Z}\pi_j/2$. The argument for the lower bound $\frac{N_j}{|V_{\text{init}}|}\ge \pi_j-\epsilon$ follows symmetrically. Hence $\sigma^*\in X_0^*(\epsilon)$ with high probability. □

We aim to show that the log-likelihood of any incorrect labeling $\sigma: V_{\text{init}}\to Z$ is strictly smaller than that of the true labeling σ^*.

We distinguish between labelings that are close to the ground truth and those that are significantly different. We formalize this using the Hamming distance up to permissible relabelings, which we term as *discrepancy*.

Definition 15 (Discrepancy)**.** The Hamming distance between two labelings $\sigma,\sigma': V_{\text{init}}\to Z$ is defined by

$$d_H(\sigma,\sigma') = \sum_{u\in V_{\text{init}}} \mathbf{1}(\sigma(u)\neq\sigma'(u)). \tag{18}$$

The discrepancy between σ and σ' is defined as the minimum Hamming distance achieved over the set of permissible relabelings $\Omega_{\pi,P}$:

$$\text{DISC}(\sigma,\sigma') = \min_{\omega\in\Omega_{\pi,P}} d_H(\omega\circ\sigma,\sigma'). \tag{19}$$

A labeling σ is considered correct if and only if $\mathrm{DISC}(\sigma, \sigma^*) = 0$. We establish the optimality of the true labeling by showing that the log-likelihood of any incorrect labeling is strictly lower than that of the ground truth.

Proposition 16 (Low Discrepancy)**.** *Suppose that Assumptions 1 and 3 hold. There exists a constant $c \in (0,1)$ such that with high probability:*

$$\forall \sigma : V_{\mathrm{init}} \to Z \text{ such that } 0 < \mathrm{DISC}(\sigma, \sigma^*) < c \log n, \text{ we have } \ell_0(G, \sigma) < \ell_0(G, \sigma^*).$$

Proposition 17 (High Discrepancy)**.** *Suppose that Assumptions 1, 2, and 3 hold. Then there exist constants $c, \epsilon \in (0,1)$, chosen as in* (24)*, such that with high probability:*

$$\forall \sigma \in X_0^*(\epsilon) \text{ such that } \mathrm{DISC}(\sigma, \sigma^*) \geq c \log n, \text{ we have } \ell_0(G, \sigma) < \ell_0(G, \sigma^*).$$

Throughout this section, we fix $t = 1/2$ for convenience. Recall the definition of ℓ and π_{wit} from (7). For any $a \neq b$ we fix some interval $I_{ab} \subset [0, r]$ with $|I_{ab}| = \ell$ such that $\sum_{w \in W_{ab}(I_{ab})} \pi_w \geq \pi_{\mathrm{wit}}$ and for any $w \in W_{ab}(I_{ab})$ and almost all $y \in I_{ab}$,

$$\phi_{1/2}(P_{aw}(y), P_{bw}(y)) < 1.$$

The existence of I_{ab} is guaranteed by Assumption 1.

By Lemma 11, for any $\epsilon > 0$ there exists some $\Phi < 1$ such that for any $w \in W_{ab}(I_{ab})$,

$$\left| \left\{ y \in I_{ab} : \phi_{1/2}(P_{aw}(y), P_{bw}(y)) < \Phi \right\} \right| \geq (1 - \epsilon)|I_{ab}| = (1 - \epsilon)\ell.$$

For $u \in B_{\mathrm{init}}$ let

$$A_{ab}(u) := \{ x \in B_{\mathrm{init}} : \frac{\|x - u\|}{(\log n)^{1/d}} \in I_{ab} \}. \tag{20}$$

By Lemma 11 and the geometry of B_{init}, there exist constants $\Phi_1 < 1$ and $\kappa_0 > 0$ such that, uniformly over $a \neq b$, $w \in W_{ab}(I_{ab})$, and $u \in B_{\mathrm{init}}$,

$$\left| \left\{ x \in A_{ab}(u) : \phi_{1/2} \left(\bar{P}_{aw}(\|x - u\|), \bar{P}_{bw}(\|x - u\|) \right) < \Phi_1 \right\} \right| \geq \kappa_0 \log n. \tag{21}$$

Let

$$\mathfrak{R} := \left\{ (\omega, a, b) : \begin{array}{l} \omega : Z \to Z \text{ is a permutation,} \\ \pi_{\omega(i)} = \pi_i \text{ for all } i \in Z, \\ P_{ab} \not\equiv P_{\omega(a)\omega(b)} \end{array} \right\}.$$

For every $(\omega, a, b) \in \mathfrak{R}$, there exists $\Phi_{ab}^{\omega} < 1$ such that

$$\left| \left\{ y \in [0, r] : \phi_{1/2} \left(P_{\omega(a)\omega(b)}(y), P_{ab}(y) \right) \leq \Phi_{ab}^{\omega} \right\} \right| > 0.$$

Since $\mathfrak{R}$ is finite, we may set

$$\Phi_2 := \max_{(\omega, a, b) \in \mathfrak{R}} \Phi_{ab}^{\omega} < 1.$$

For $(\omega, a, b) \in \mathfrak{R}$, let

$$J_{ab}^{\omega} := \left\{ y \in [0, r] : \phi_{1/2} \left(P_{\omega(a)\omega(b)}(y), P_{ab}(y) \right) \leq \Phi_2 \right\}.$$

Then $|J_{ab}^{\omega}| > 0$ for every $(\omega, a, b) \in \mathfrak{R}$. For measurable $J \subseteq [0, r]$ let

$$A_J(u) := \left\{ x \in B_{\mathrm{init}} : \frac{\|x - u\|}{(\log n)^{1/d}} \in J \right\}.$$

There is a deterministic constant $\kappa > 0$, uniform over all $(\omega, a, b) \in \mathfrak{R}$ and all $u \in B_{\mathrm{init}}$, such that

$$|A_{J_{ab}^{\omega}}(u)| \geq \kappa \log n.$$

Indeed, we may choose

$$\kappa = \frac{d \nu_d}{2^d} \min_{(\omega, a, b) \in \mathfrak{R}} \int_{J_{ab}^{\omega}} s^{d-1} \mathrm{d}s > 0.$$

Now let $\pi_{\min} := \min_{a \in Z} \pi_a$, $m_0 := \lambda \pi_{\min} \kappa$, and $\eta := m_0/2 > 0$. Then

$$\mathcal{G}_{\text{shell}} := \bigcap_{(\omega,a,b)\in\mathfrak{R}} \bigcap_{u \in V_{\text{init}}} \left\{ \left| (V_b \setminus \{u\}) \cap A_{J^{\omega}_{ab}}(u) \right| \geq \eta \log n \right\}. \tag{22}$$

Then,

$$\begin{aligned}\mathbb{P}(\mathcal{G}^c_{\text{shell}}) &\leq \mathbb{E}\left[\sum_{u \in V_{\text{init}}} \sum_{(\omega,a,b)\in\mathfrak{R}} \mathbf{1}\left\{ \left| (V_b \setminus \{u\}) \cap A_{J^{\omega}_{ab}}(u) \right| < \eta \log n \right\} \right] \\ &\leq \lambda |B_{\text{init}}| |\mathfrak{R}|\, n^{-m_0/8} = o(1),\end{aligned}$$

where we have used the Mecke equation (Proposition 8) to obtain the second inequality.

Fix $c_-, c_+ > 0$ such that the event

$$\mathcal{G}_0 := \{ c_- \log n \leq |V_{\text{init}}| \leq c_+ \log n \} \tag{23}$$

occurs with high probability. Fix $c, \epsilon > 0$ such that

$$c < \frac{\lambda \pi_{\text{wit}} \kappa_0 \log(1/\Phi_1)}{4\rho}, \quad \epsilon \leq \min\left\{ \frac{\pi_{\min}}{2}, \min_{i,j \in Z, \pi_i \neq \pi_j} \frac{|\pi_i - \pi_j|}{3}, \frac{c}{(k-1)c_+}, \frac{k\eta}{2(k-1)c_+} \right\}. \tag{24}$$

Proof of Proposition 13. Consider the estimated labeling $\hat{\sigma}$ that maximizes the restricted MLE, i.e.

$$\hat{\sigma} \in \operatorname*{arg\,max}_{\sigma \in X_0^*(\epsilon)} \ell_0(G, \sigma). \tag{25}$$

The probability that $\hat{\sigma}$ does not achieve exact recovery (i.e. that $\text{DISC}(\hat{\sigma}, \sigma^*) \neq 0$) can be bounded by

$$\begin{aligned}&\mathbb{P}\Big(\bigcup_{\substack{\sigma \in X_0^*(\epsilon) \\ \text{DISC}(\sigma,\sigma^*)\neq 0}} \{\ell_0(G,\sigma) \geq \ell_0(G,\sigma^*)\} \Big) \\ &\leq \mathbb{P}\Big(\bigcup_{\substack{\sigma \in X_0^*(\epsilon) \\ \text{DISC}(\sigma,\sigma^*)\neq 0}} \{\ell_0(G,\sigma) \geq \ell_0(G,\sigma^*)\} \cap \{\sigma^* \in X_0^*(\epsilon)\} \Big) + \mathbb{P}(\sigma^* \notin X_0^*(\epsilon)).\end{aligned} \tag{26}$$

Propositions 16 and 17 yield that the first probability tends to zero, and Lemma 14 gives that the second probability tends to zero. This finishes the proof. □

We finally transfer this result to the MAP estimator used in Algorithm 1. For a labeling σ, let

$$[\sigma] := \{\omega \circ \sigma : \omega \in \Omega_{\pi,P}\}$$

denote its equivalence class under permissible relabelings. By the definition of a permissible relabeling, all labelings in the same equivalence class have the same posterior probability.

Let $\mathcal{D}_{\text{init}}$ denote the observed vertex locations in V_{init} together with all edge weights between pairs of vertices in V_{init}. Let

$$Q(\mathcal{D}_{\text{init}}) := \mathbb{P}\big([\sigma^*] = [\bar{\sigma}] \,\big|\, \mathcal{D}_{\text{init}}\big)$$

where $\bar{\sigma}$ is the restricted MLE. By Proposition 13,

$$\mathbb{E}[1 - Q(\mathcal{D}_{\text{init}})] = \mathbb{P}\big([\bar{\sigma}] \neq [\sigma^*]\big) = o(1),$$

and hence $Q(\mathcal{D}_{\text{init}}) = 1 - o_{\mathbb{P}}(1)$.

Since every equivalence class contains at most $k!$ labelings, on the event

$$Q(\mathcal{D}_{\text{init}}) > \frac{k!}{k!+1},$$

each labeling in $[\bar{\sigma}]$ has posterior probability greater than $1/(k!+1)$, while the total posterior probability of all labelings outside $[\bar{\sigma}]$ is less than $1/(k!+1)$. Consequently, every MAP labeling belongs to $[\bar{\sigma}]$. It follows from Proposition 13 that the MAP estimator achieves exact recovery on V_{init} with probability $1 - o(1)$.

Now it remains to prove Propositions 16 and 17.

Proof of Proposition 16. Fix some $\sigma : V_{\text{init}} \to Z$ with $\text{DISC}(\sigma, \sigma^*) > 0$. Without loss of generality, after a permissible relabeling, we assume that $d_H(\sigma, \sigma^*) = \text{DISC}(\sigma, \sigma^*)$. Let $d_{uv} := \|u - v\|$. The log-likelihood difference between σ and the ground truth labeling σ^* can be expressed as

$$\begin{aligned}
\ell_0(G,\sigma) - \ell_0(G,\sigma^*) &= \sum_{\substack{u,v\in V_{\text{init}}\\ u\neq v}} \log \frac{\bar{p}_{\sigma(u)\sigma(v)}(x_{uv}; d_{uv})}{\bar{p}_{\sigma^*(u)\sigma^*(v)}(x_{uv}; d_{uv})} \\
&= \sum_{\substack{u,v\in V_{\text{init}}\\ u\neq v}} \log \frac{\bar{p}_{\sigma(u)\sigma^*(v)}(x_{uv}; d_{uv})}{\bar{p}_{\sigma^*(u)\sigma^*(v)}(x_{uv}; d_{uv})} + \sum_{\substack{u,v\in V_{\text{init}}\\ u\neq v}} \log \frac{\bar{p}_{\sigma(u)\sigma(v)}(x_{uv}; d_{uv})}{\bar{p}_{\sigma(u)\sigma^*(v)}(x_{uv}; d_{uv})} \\
&= \sum_{\substack{u,v\in V_{\text{init}}\\ u\neq v,\ \sigma(u)\neq\sigma^*(u)}} \log \frac{\bar{p}_{\sigma(u)\sigma^*(v)}(x_{uv}; d_{uv})}{\bar{p}_{\sigma^*(u)\sigma^*(v)}(x_{uv}; d_{uv})} + \sum_{\substack{u,v\in V_{\text{init}}\\ u\neq v,\ \sigma(v)\neq\sigma^*(v)}} \log \frac{\bar{p}_{\sigma(u)\sigma(v)}(x_{uv}; d_{uv})}{\bar{p}_{\sigma(u)\sigma^*(v)}(x_{uv}; d_{uv})}.
\end{aligned}$$

We decompose the last expression into $A(\sigma) + B(\sigma)$ with

$$A(\sigma) := \sum_{\substack{u\in V_{\text{init}}\\ \sigma(u)\neq\sigma^*(u)}} \sum_{\substack{v\in V_{\text{init}}\\ v\neq u}} \log\left(\frac{\bar{p}_{\sigma(u)\sigma^*(v)}(x_{uv}; \|u-v\|)}{\bar{p}_{\sigma^*(u)\sigma^*(v)}(x_{uv}; \|u-v\|)}\right).$$

Since there are exactly $d_H(\sigma, \sigma^*)$ points with $\sigma(u) \neq \sigma^*(u)$, it holds that

$$\begin{aligned}
&\mathbb{P}\Big(\bigcup_{\sigma: V_{\text{init}}\to Z} \{A(\sigma) > -d_H(\sigma,\sigma^*) c_1 \log n\}\Big) \\
&\leq \mathbb{P}\Big(\bigcup_{u\in V_{\text{init}}} \bigcup_{b\neq\sigma^*(u)} \Big\{\sum_{\substack{v\in V_{\text{init}}\\ v\neq u}} \log\Big(\frac{\bar{p}_{b\sigma^*(v)}(x_{uv}; \|u-v\|)}{\bar{p}_{\sigma^*(u)\sigma^*(v)}(x_{uv}; \|u-v\|)}\Big) > -c_1 \log n\Big\}\Big) \\
&\leq \mathbb{E}\Big[\sum_{u\in V_{\text{init}}} \sum_{b\neq\sigma^*(u)} \mathbf{1}\Big\{\sum_{\substack{v\in V_{\text{init}}\\ v\neq u}} \log\Big(\frac{\bar{p}_{b\sigma^*(v)}(x_{uv}; \|u-v\|)}{\bar{p}_{\sigma^*(u)\sigma^*(v)}(x_{uv}; \|u-v\|)}\Big) > -c_1 \log n\Big\}\Big].
\end{aligned}$$

By the Mecke equation (Proposition 8), the above is given by

$$\lambda \int_{B_{\text{init}}} \sum_{a\in Z} \pi_a \sum_{b\in Z\setminus\{a\}} \mathbb{P}\Big(\sum_{v\in(V_{\text{init}}\cup\{u\})\setminus\{u\}} \log\Big(\frac{\bar{p}_{b\sigma^*(v)}(x_{uv}; \|u-v\|)}{\bar{p}_{a\sigma^*(v)}(x_{uv}; \|u-v\|)}\Big) > -c_1 \log n\Big) \mathrm{d}u. \quad (27)$$

Fix $a, b \in Z$ with $a \neq b$ and let

$$S_u^{ab} = \sum_{v\in(V_{\text{init}}\cup\{u\})\setminus\{u\}} \log\Big(\frac{\bar{p}_{b\sigma^*(v)}(x_{uv}; \|u-v\|)}{\bar{p}_{a\sigma^*(v)}(x_{uv}; \|u-v\|)}\Big)$$

be the log-likelihood contribution of u.

For fixed $a \neq b$, define a set of informative marked points

$$E_{ab}(u) := \{(x,w) : x \in A_{ab}(u), w \in W_{ab}(I_{ab}), \phi_{1/2}(\bar{P}_{aw}(\|u-x\|), \bar{P}_{bw}(\|u-x\|)) < \Phi_1\},$$

where $A_{ab}(u)$ is defined at (20). By (21), uniformly in u,

$$\sum_{w\in W_{ab}(I_{ab})} \pi_w |\{x : (x,w) \in E_{ab}(u)\}| \geq \kappa_0 \pi_{\text{wit}} \log n.$$

Therefore the number N_u^{ab} of Poisson points falling in $E_{ab}(u)$ satisfies

$$N_u^{ab} \sim \text{Poisson}(\mu_u^{ab}), \qquad \mu_u^{ab} \geq \kappa_0 \lambda \pi_{\text{wit}} \log n.$$

Hence, by a Poisson Chernoff bound (Lemma 9),

$$\mathbb{P}\left(N_u^{ab} \leq \frac{2}{3}\mu_u^{ab}\right) \leq \exp\left(-\frac{1}{18}\mu_u^{ab}\right) \leq n^{-\lambda\pi_{\text{wit}}\kappa_0/18}. \quad (28)$$

Let

$$c_1 := \frac{1}{4}\lambda\pi_{\mathrm{wit}}\kappa_0 \log(1/\Phi_1) > 0.$$

We bound the probability that S_u^{ab} exceeds $-c_1 \log n$ by distinguishing according to the value of N_u^{ab}. Since $t = 1/2$, Markov's inequality gives

$$\begin{aligned}\mathbb{P}(S_u^{ab} \geq -c_1 \log n) &\leq \mathbb{P}\left(\{S_u^{ab} \geq -c_1 \log n\} \cap \left\{N_u^{ab} > \frac{2}{3}\mu_u^{ab}\right\}\right) \\ &\quad + \mathbb{P}\left(N_u^{ab} \leq \frac{2}{3}\mu_u^{ab}\right) \\ &\leq n^{c_1/2}\mathbb{E}\left[e^{S_u^{ab}/2}\mathbf{1}\left\{N_u^{ab} > \frac{2}{3}\mu_u^{ab}\right\}\right] + \mathbb{P}\left(N_u^{ab} \leq \frac{2}{3}\mu_u^{ab}\right). \end{aligned} \tag{29}$$

Conditioning on V and σ^*, we have

$$\mathbb{E}\left[e^{S_u^{ab}/2} \,\middle|\, V, \sigma^*\right] = \prod_{v \in V_{\mathrm{init}}} \phi_{1/2}\left(\bar{P}_{b\sigma^*(v)}(\|u-v\|), \bar{P}_{a\sigma^*(v)}(\|u-v\|)\right) \leq \Phi_1^{N_u^{ab}}.$$

Consequently,

$$\begin{aligned}\mathbb{P}(S_u^{ab} \geq -c_1 \log n) &\leq n^{c_1/2}\Phi_1^{\frac{2}{3}\mu_u^{ab}} + n^{-\lambda\pi_{\mathrm{wit}}\kappa_0/18} \\ &\leq n^{c_1/2} n^{-\frac{2}{3}\lambda\pi_{\mathrm{wit}}\kappa_0 \log(1/\Phi_1)} + n^{-\lambda\pi_{\mathrm{wit}}\kappa_0/18} \\ &= n^{-\frac{13}{24}\lambda\pi_{\mathrm{wit}}\kappa_0 \log(1/\Phi_1)} + n^{-\lambda\pi_{\mathrm{wit}}\kappa_0/18} \\ &= o(1/\log n).\end{aligned}$$

Together with (27), this gives

$$\mathbb{P}\left(\bigcap_{\sigma: V_{\mathrm{init}} \to Z} \{A(\sigma) \leq -d_H(\sigma, \sigma^*) c_1 \log n\}\right) = 1 - o(1). \tag{30}$$

It remains to bound $B(\sigma)$. Following the decomposition of the term $B(\sigma)$ in the proof of Gaudio & Jin (2026), we split

$$B(\sigma) = B_1(\sigma) + B_2(\sigma), \tag{31}$$

where

$$B_1(\sigma) = \sum_{\substack{v \in V_{\mathrm{init}} \\ \sigma(v) \neq \sigma^*(v)}} \sum_{\substack{u \in V_{\mathrm{init}} \\ u \neq v}} \log \frac{\bar{p}_{\sigma^*(u),\sigma(v)}(x_{uv}; \|u-v\|)}{\bar{p}_{\sigma^*(u),\sigma^*(v)}(x_{uv}; \|u-v\|)} \tag{32}$$

which is identical to $A(\sigma)$ and

$$B_2(\sigma) = \sum_{\substack{u,v \in V_{\mathrm{init}},\, u \neq v \\ \sigma(u) \neq \sigma^*(u),\, \sigma(v) \neq \sigma^*(v)}} \log \frac{\bar{p}_{\sigma(u),\sigma(v)}(x_{uv}; \|u-v\|)\, \bar{p}_{\sigma^*(u),\sigma^*(v)}(x_{uv}; \|u-v\|)}{\bar{p}_{\sigma(u),\sigma^*(v)}(x_{uv}; \|u-v\|)\, \bar{p}_{\sigma^*(u),\sigma(v)}(x_{uv}; \|u-v\|)}. \tag{33}$$

By Assumption 3, each summand of $B_2(\sigma)$ is bounded by 2ρ, the number of summands is at most $d_H(\sigma, \sigma^*)^2$. Hence $B_2(\sigma) \leq 2\rho\, d_H(\sigma, \sigma^*)^2$. Combining these bounds, with high probability

$$\ell_0(G, \sigma) - \ell_0(G, \sigma^*) \;\leq\; -2\, d_H(\sigma, \sigma^*)\, c_1 \log n + 2\rho\, d_H(\sigma, \sigma^*)^2 \tag{34}$$

for every σ. When $0 < d_H(\sigma, \sigma^*) < c \log n$ with $c < c_1/\rho$, the right-hand side is strictly negative. This concludes the proof of exact recovery in the initial cube in the low-discrepancy case. Thus $\ell_0(G, \sigma) - \ell_0(G, \sigma^*) < 0$, with high probability, when $0 < \mathrm{DISC}(\sigma, \sigma^*) < c \log n$ with $c < \frac{c_1}{\rho}$. □

Lemma 18 (Geometric split)**.** *Fix $s \in (0, r]$ and $\delta > 0$. There exists $\gamma = \gamma(s, \delta) > 0$ such that, with probability $1 - o(1)$, the following holds simultaneously for every $a \in Z$ and every labeling*

$$\sigma : V \cap B_{\mathrm{init}} \to Z.$$

Suppose that there exist $b, b' \in Z$, with $b \neq b'$, such that

$$|\{u \in V_a \cap B_{\text{init}} : \sigma(u) = b\}| \geq \delta \log n$$

and

$$|\{u \in V_a \cap B_{\text{init}} : \sigma(u) = b'\}| \geq \delta \log n.$$

Then there exists a collection $\mathcal{E}$ of unordered pairs $\{u, v\} \subseteq V_a \cap B_{\text{init}}$ such that

$$|\mathcal{E}| \geq \gamma (\log n)^2,$$

and, for every $\{u, v\} \in \mathcal{E}$,

$$\sigma(u) \neq \sigma(v) \qquad \textit{and} \qquad \|u - v\| \leq s(\log n)^{1/d}.$$

Proof. Rescale B_{init} by $(\log n)^{-1/d}$, so that it becomes the fixed cube $[-r, r]^d$.

Choose an integer $m \geq 1$ sufficiently large that, if $[-r, r]^d$ is partitioned into m^d congruent cubes, then any two points lying either in the same cube or in two face-adjacent cubes are at Euclidean distance at most s.

Let

$$\mathcal{C} = \{C_1, \ldots, C_M\}, \qquad M = m^d,$$

denote the corresponding partition of B_{init}, scaled back by $(\log n)^{1/d}$.

Since the points of community a form a Poisson point process of intensity $\lambda \pi_a$, there exists $\alpha > 0$ such that, with probability $1 - o(1)$,

$$|V_a \cap C| \geq \alpha \log n \tag{35}$$

simultaneously for every $a \in Z$ and $C \in \mathcal{C}$. Indeed, there are only $kM = O(1)$ such random variables, and each is Poisson with mean of order $\log n$.

Work on the event (35). Set

$$\tau := \min \left\{ \frac{\delta}{2M}, \frac{\alpha}{2k} \right\}.$$

For $C \in \mathcal{C}$ and $j \in Z$, call j *heavy in* C if

$$|\{u \in V_a \cap C : \sigma(u) = j\}| \geq \tau \log n.$$

By (35) and the pigeonhole principle, every cube has at least one heavy label. Moreover, since

$$|\{u \in V_a \cap B_{\text{init}} : \sigma(u) = b\}| \geq \delta \log n,$$

there exists a cube in which b is heavy. The same holds for b'.

Suppose first that some cube C contains two distinct heavy labels $j \neq j'$. Then there are at least $\tau^2 (\log n)^2$ pairs $\{u, v\} \subseteq V_a \cap C$ with $\sigma(u) = j$ and $\sigma(v) = j'$. By the choice of the partition,

$$\|u - v\| \leq s(\log n)^{1/d},$$

and the result follows.

It remains to consider the case in which every cube has a unique heavy label. Assign to each cube its unique heavy label. The face-adjacency graph of the cubes is connected. Since some cube has heavy label b and some other cube has heavy label b', there must exist two face-adjacent cubes C, C' whose heavy labels, say j and j', are distinct.

Consequently, there are at least $\tau^2 (\log n)^2$ pairs $u \in V_a \cap C$, $v \in V_a \cap C'$ such that

$$\sigma(u) = j, \qquad \sigma(v) = j'.$$

Again, by construction,

$$\|u - v\| \leq s(\log n)^{1/d}.$$

Thus, the assertion holds with $\gamma := \tau^2$. □

Proof of Proposition 17. Recall the definitions of c_- and c_+ from (23) and let

$$\widetilde{\mathcal{G}}_0 := \{\sigma^* \in X_0^*(\epsilon),\ c_- \log n \le |V_{\text{init}}| \le c_+ \log n\},$$

and let $\mathcal{G}_{\text{split}}$ be the high-probability event given by Lemma 18 for

$$s = r_{\text{init}}, \quad \delta = \frac{\epsilon c_-}{k}.$$

We work on the event $\mathcal{G} := \widetilde{\mathcal{G}}_0 \cap \mathcal{G}_{\text{split}} \cap \mathcal{G}_{\text{shell}}$. Then $\mathbb{P}(\mathcal{G}) = 1 - o(1)$. Recall the definitions of $c > 0$ and let $\epsilon > 0$ from (24). We consider a labeling $\sigma \in X_0^*(\epsilon)$ satisfying

$$\text{DISC}(\sigma, \sigma^*) \ge c \log n.$$

For $i, j \in Z$, let

$$N_{ij} = \{u \in V_{\text{init}} : \sigma^*(u) = i,\ \sigma(u) = j\}, \qquad n_{ij} = |N_{ij}|.$$

Since $P_{ij} = P_{ji}$ for all $i, j \in Z$, we have,

$$\ell_0(G, \sigma) - \ell_0(G, \sigma^*) = 2 \sum_{\{u,v\} \subseteq V_{\text{init}}} \log\left(\frac{\bar{p}_{\sigma(u),\sigma(v)}(x_{uv}; \|u - v\|)}{\bar{p}_{\sigma^*(u),\sigma^*(v)}(x_{uv}; \|u - v\|)}\right)$$

Conditionally on V and σ^*, Markov's inequality gives

$$\begin{aligned}
&\mathbb{P}\left(\ell_0(G, \sigma) - \ell_0(G, \sigma^*) \ge 0 \,|\, V, \sigma^*\right) \\
&\quad \le \mathbb{E}\left[\exp\left\{\frac{1}{4}\big(\ell_0(G, \sigma) - \ell_0(G, \sigma^*)\big)\right\} \,\middle|\, V, \sigma^*\right] \\
&\quad = \prod_{\{u,v\} \subseteq V_{\text{init}}} \phi_{1/2}\left(\bar{P}_{\sigma(u)\sigma(v)}(\|u - v\|), \bar{P}_{\sigma^*(u)\sigma^*(v)}(\|u - v\|)\right). \qquad (36)
\end{aligned}$$

Since the vertex locations have an absolutely continuous distribution, almost surely none of the finitely many interpoint distances in V_{init} belongs to any exceptional null set in Assumption 2. We now distinguish two cases.

Case 1. Suppose that there exist $a, b, b' \in Z$, with $b \ne b'$, such that

$$n_{ab}, n_{ab'} \ge \frac{\epsilon}{k}|V_{\text{init}}|.$$

On the event $|V_{\text{init}}| \ge c_- \log n$, Lemma 18, applied with

$$s = r_{\text{init}}, \qquad \delta = \frac{\epsilon c_-}{k},$$

gives a set $\mathcal{E}$ of at least $\gamma(\log n)^2$ unordered pairs $u, v \in V_{\text{init}}$ such that

$$\sigma^*(u) = \sigma^*(v), \qquad \sigma(u) \ne \sigma(v), \qquad \|u - v\| \le r_{\text{init}}(\log n)^{1/d}$$

for every $\{u, v\} \in \mathcal{E}$.

Assumption 2 therefore yields

$$\phi_{1/2}\left(\bar{P}_{\sigma(u)\sigma(v)}(\|u - v\|), \bar{P}_{\sigma^*(u)\sigma^*(v)}(\|u - v\|)\right) \le \Phi_{\text{init}}$$

for every $\{u, v\} \in \mathcal{E}$. Hence

$$\mathbb{P}\left(\ell_0(G, \sigma) - \ell_0(G, \sigma^*) \ge 0\right) \le \Phi_{\text{init}}^{\gamma(\log n)^2} = C_1^{(\log n)^2}$$

for some $C_1 \in (0, 1)$.

Case 2. Suppose that for every $a \in Z$ there is at most one $b \in Z$ such that

$$n_{ab} \ge \frac{\epsilon}{k}|V_{\text{init}}|.$$

We claim that for every a there is in fact exactly one such b. Indeed, if no such b existed for some a, then

$$\sum_{b \in Z} n_{ab} < \epsilon |V_{\text{init}}|.$$

On the other hand, since $\sigma^* \in X_0^*(\epsilon)$,

$$\sum_{b \in Z} n_{ab} = |\{u : \sigma^*(u) = a\}| \geq (\pi_a - \epsilon)|V_{\text{init}}|.$$

Since $\epsilon \leq \pi_{\min}/2$, this is at least $\epsilon |V_{\text{init}}|$, which is a contradiction.

Hence there is a unique map $\omega : Z \to Z$ such that

$$n_{a\omega(a)} \geq \frac{\epsilon}{k}|V_{\text{init}}| \qquad \text{for every } a \in Z.$$

We next show that ω is a permutation. Suppose that some $b \in Z$ is not in the image of ω. Then, by the assumption of Case 2,

$$n_{ab} < \frac{\epsilon}{k}|V_{\text{init}}| \qquad \text{for every } a \in Z.$$

Consequently,

$$|\{u \in V_{\text{init}} : \sigma(u) = b\}| = \sum_{a \in Z} n_{ab} < \epsilon |V_{\text{init}}|.$$

But $\sigma \in X_0^*(\epsilon)$ implies

$$|\{u : \sigma(u) = b\}| \geq (\pi_b - \epsilon)|V_{\text{init}}|,$$

which is at least $\epsilon |V_{\text{init}}|$ since $\epsilon \leq \pi_{\min}/2$. This is a contradiction. Thus ω is bijective.

We now show that ω cannot be a permissible relabeling. If it were permissible, then ω^{-1} would also be permissible and

$$\text{DISC}(\sigma, \sigma^*) \leq d_H(\omega^{-1} \circ \sigma, \sigma^*) = \sum_{a \in Z} \sum_{b \neq \omega(a)} n_{ab}.$$

By the assumption of Case 2,

$$\sum_{a \in Z} \sum_{b \neq \omega(a)} n_{ab} < k(k-1)\frac{\epsilon}{k}|V_{\text{init}}| = (k-1)\epsilon |V_{\text{init}}|.$$

Using (24) gives

$$\text{DISC}(\sigma, \sigma^*) < c \log n,$$

contradicting the assumption on σ. Hence ω is not permissible.

Following the argument in the proof of Proposition D.6 Gaudio et al. (2025), we next establish that $\pi_a = \pi_{\omega(a)}$ for every $a \in Z$. Suppose that $\pi_{\omega(a)} \neq \pi_a$ for some $a \in Z$. Since ω is a permutation, there must then exist some a such that $\pi_{\omega(a)} < \pi_a$. Since $\sum_{b \in Z} n_{ab} \geq (\pi_a - \epsilon)|V_{\text{init}}|$ with high probability, and $n_{ab} < \frac{\epsilon}{k}|V_{\text{init}}|$ for any $b \neq \omega(a)$, we have

$$n_{a\omega(a)} > (\pi_a - \epsilon)|V_{\text{init}}| - (k-1)|V_{\text{init}}|\frac{\epsilon}{k} > (\pi_a - 2\epsilon)|V_{\text{init}}|.$$

Together with the definition of $X_0^*(\epsilon)$, this gives

$$(\pi_a - 2\epsilon)|V_{\text{init}}| < \sum_{b \in Z} n_{b\omega(a)} \leq (\pi_{\omega(a)} + \epsilon)|V_{\text{init}}|.$$

Rearranging terms yields $\epsilon > (\pi_a - \pi_{\omega(a)})/3$. This contradicts the definition of ϵ. Hence, for all $a \in Z$ we must have $\pi_{\omega(a)} \geq \pi_a$, which implies $\pi_{\omega(a)} = \pi_a$, since the $\{\pi_a\}$ sum to 1.

Since ω is not permissible, there exist $a, b \in Z$ such that

$$P_{ab} \not\equiv P_{\omega(a)\omega(b)}.$$

Hence $(\omega, a, b) \in \mathfrak{R}$. On $\mathcal{G}_{\text{shell}}$, for every $u \in N_{a\omega(a)}$,

$$\left|(V_b \setminus \{u\}) \cap A_{J^{\omega}_{ab}}(u)\right| \geq \eta \log n.$$

On the other hand, by the assumption of Case 2,

$$|(V_b \cap V_{\text{init}}) \setminus N_{b\omega(b)}| = \sum_{j \neq \omega(b)} n_{bj} < \frac{k-1}{k}\epsilon|V_{\text{init}}| \leq \frac{k-1}{k}\epsilon c_+ \log n.$$

Thus,

$$|(N_{b\omega(b)} \setminus \{u\}) \cap A_{J^{\omega}_{ab}}(u)| \geq \frac{\eta}{2}\log n.$$

On $\mathcal{G}_0$,

$$n_{a\omega(a)} \geq \frac{\epsilon}{k}|V_{\text{init}}| \geq \frac{\epsilon c_-}{k}\log n.$$

Hence, with $\gamma_2 := (\epsilon c_- \eta)/4k$, there are at least $\gamma_2(\log n)^2$ unordered pairs $u, v \in V_{\text{init}}$ satisfying

$$\phi_{1/2}\left(\bar{P}_{ab}(\|u-v\|), \bar{P}_{\omega(a)\omega(b)}(\|u-v\|)\right) \leq \Phi_2.$$

Therefore,

$$\mathbb{P}\left(\ell_0(G,\sigma) \geq \ell_0(G,\sigma^*) \,|\, V, \sigma^*\right) \leq (\Phi_2^{\gamma_2})^{(\log n)^2}.$$

Let $\widetilde{C} := \max\{C_1, \Phi_2^{\gamma_2}\} < 1$. Consequently,

$$\begin{aligned}&\mathbb{P}\left(\exists\, \sigma \in X_0^*(\epsilon) : \text{DISC}(\sigma,\sigma^*) \geq c\log n,\ \ell_0(G,\sigma) \geq \ell_0(G,\sigma^*)\right)\\ &\quad\leq \mathbb{P}(\mathcal{G}^c) + k^{c_+\log n}\widetilde{C}^{(\log n)^2} = o(1).\end{aligned}$$

□

H PROPAGATION PROOF

H.1 BLOCK STRUCTURE

We prepare the propagation proof of Theorem 3 by providing details on the discretization of the space $\mathcal{S}_{d,n}$ into blocks. Recall the definitions of π_{wit} and ℓ from (7).

Lemma 19 ((Gaudio et al., 2025, Section C)). *Let $d = 1$ and suppose that $\lambda\pi_{\text{wit}}\ell > 1$. Denote by K_1 the number of blocks ℓ-close to the left (or the right) of length $\ell\chi\log n$. Then, we have $\lambda\pi_{\text{wit}}K_1\chi\ell > 1$ for all χ satisfying* (8)*.*

For $d \geq 2$ suppose that $\lambda\pi_{\text{wit}}\nu_d\ell^d > 1$. Then there exists $\chi_0 > 0$ such that for all $0 < \chi < \chi_0$, (9) *holds. Moreover, let K_d be the number of blocks that are ℓ-close to a given block, excluding the block itself. Since the torus is translation invariant and the partition is regular, it depends only on χ and d. For all χ verifying the condition in* (9)*, we have $\lambda\pi_{\text{wit}}K_d\chi\ell^d > 1$.*

Proof. For $d = 1$, $K_1 = \lfloor 1/\chi \rfloor - 1$, then $\chi K_1 \geq 1 - 2\chi$ and from (8), we obtain the result.

For $d \geq 2$, note that $h(x) := \lambda\pi_{\text{wit}}\ell^d\left(\nu_d\left(1 - \frac{3\sqrt{d}}{2}x^{1/d}\right)^d - x\right)$ is decreasing, continuous, and $\lim_{x\downarrow 0} h(x) = \lambda\pi_{\text{wit}}\ell^d\nu_d > 1$. Hence, there exists $\chi_0' > 0$ such that for all $0 < \chi < \chi_0'$, $h(\chi) > 1$. Hence, we take $\chi_0 < \min\left\{\chi_0', \left(\frac{2}{3\sqrt{d}}\right)^d\right\}$. We define $\ell_d := \ell - \frac{3}{2}\sqrt{d}\ell\chi^{1/d}$ the proximity radius reduced by one and a half block diagonals. For a block $B \in \mathcal{P}_\chi$, let $c(B)$ denote its center and define $C(B) := \mathcal{B}\left(c(B), \ell_d(\log n)^{1/d}\right)$. By second condition in (9), $\ell_d > 0$. We define the set of blocks B' such that B and B' are ℓ-close by $U(B) := \bigcup_{\substack{B' \sim B \\ B' \neq B}} B'$.

Note that $C(B) \subset U(B) \cup B$. Hence,

$$K_d\chi\ell^d\log n = |U(B)| \geq |C(B)| - |B| \geq \ell^d\left(\nu_d\left(1 - \frac{3\sqrt{d}}{2}\chi^{1/d}\right)^d - \chi\right)\log n.$$

Then from (9), $\lambda\pi_{\text{wit}}K_d\chi\ell^d \geq \lambda\pi_{\text{wit}}\ell^d\left(\nu_d(1 - \frac{3\sqrt{d}}{2}\chi^{1/d})^d - \chi\right) > 1$. □

Proof of Proposition 7. The "if" direction is Lemma 19; we prove the converse.

Let $d = 1$. Condition (8) reads $0 < \chi < \frac{1}{2}\big(1 - (\lambda\pi_{\mathrm{wit}}\ell)^{-1}\big)$, whose right-hand side is positive if and only if $\lambda\pi_{\mathrm{wit}}\ell > 1$. Since $\nu_1 = 2 = c_1$, this is exactly $\lambda\nu_1\ell\,\pi_{\mathrm{wit}} > c_1$.

Let $d \geq 2$ and put

$$h(\chi) := \lambda\pi_{\mathrm{wit}}\ell^d\Big(\nu_d\big(1 - \tfrac{3\sqrt{d}}{2}\chi^{1/d}\big)^d - \chi\Big), \qquad 0 < \chi < \Big(\tfrac{2}{3\sqrt{d}}\Big)^d,$$

the range on which the second requirement of (9) holds. As in the proof of Lemma 19, h is continuous and strictly decreasing there, with $\lim_{\chi\downarrow 0} h(\chi) = \lambda\pi_{\mathrm{wit}}\nu_d\ell^d$. Hence some $\chi > 0$ satisfies (9) if and only if $\lambda\nu_d\ell^d\pi_{\mathrm{wit}} > 1 = c_d$. Given such a χ, both requirements of (10) hold after decreasing χ, as noted after (10), so admissibility is unaffected.

In both cases an admissible χ exists for a given $\big(\ell, \{I_{ab}\}\big)$ if and only if $\lambda\nu_d\ell^d\pi_{\mathrm{wit}} > c_d$. Taking the supremum over $\ell \in (0, r]$ and over families of intervals of common length ℓ turns the left-hand side into the quantity appearing in Assumption 1, which proves the equivalence. □

For $w \in Z$ let $V_w := \{v \in V : \sigma^*(v) = w\}$ be the process of points labeled w under σ^*. For $B \subseteq \mathcal{S}_{d,n}$ let $V_w(B) := |V_w \cap B|$.

Definition 20 (Spread block)**.** A block B is called δ-spread if

$$V(B) \geq \delta \log n$$

and

$$\min_{a \neq b} \sum_{w \in W_{ab}(I_{ab})} V_w(B) \geq \frac{3}{4}\delta\pi_{\mathrm{wit}} \log n.$$

Definition 21 (Cluster of blocks, (Gaudio et al., 2025, Definition C.2))**.** Two blocks are adjacent if they share an edge or a corner. We say that a set of blocks $\mathcal{B}$ is a cluster if for every $B, B' \in \mathcal{B}$, there is a path of blocks of the form $(B = B_{j_1}, B_{j_2}, \ldots, B_{j_m} = B')$, where $B_{j_k} \in \mathcal{B}$ for $k \in [m]$ and $B_{j_k}, B_{j_{k+1}}$ are adjacent.

The following lemma implies that with high probability for every block $B \in \mathcal{P}_\chi$ there exists a δ-spread block B' such that B and B' are ℓ-close. It generalizes (Gaudio et al., 2025, Lemma C.8).

Lemma 22. *Suppose that $\lambda\pi_{\mathrm{wit}}\ell > 1$ if $d = 1$, and $\lambda\pi_{\mathrm{wit}}\nu_d\ell^d > 1$ if $d \geq 2$. Let Y_δ be the size of the largest cluster of blocks that are not δ-spread. Then*

$$\mathbb{P}(Y_\delta < K_d) = 1 - o(1),$$

where K_d and δ are given at (12) *and* (13)*.*

Proof. Let $\mathcal{S}_{K_d}$ be the (finite) set of cluster shapes of K_d blocks. We have $\Theta(n/\log n)$ blocks, hence there are $O(n/\log n)$ clusters of K_d blocks to consider. Let $(B_{i_1}, \ldots, B_{i_{K_d}})$ be a cluster of K_d blocks. It suffices to show that the probability that all blocks in the cluster are not δ-spread is $o((\log n)/n)$.

If B is not δ-spread, then either $V(B) < \delta \log n$, or there exist $a \neq b$ such that

$$\sum_{w \in W_{ab}(I_{ab})} V_w(B) < \frac{3}{4}\delta\pi_{\mathrm{wit}} \log n.$$

Consequently,

$$\mathbb{P}(B \text{ is not } \delta\text{-spread}) \leq \mathbb{P}(V(B) < \delta \log n) + \sum_{a<b} \mathbb{P}\left(\sum_{w \in W_{ab}(I_{ab})} V_w(B) < \frac{3}{4}\delta\pi_{\mathrm{wit}} \log n\right).$$

Hence, by independence and by translation invariance,

$$\begin{aligned}
&\mathbb{P}\Big(\bigcap_{j=1}^{K_d} \{B_{i_j} \text{ is not } \delta\text{-spread}\}\Big) \\
&\quad \leq \Big(\mathbb{P}\left(V(B) < \delta \log n\right) + \sum_{a<b} \mathbb{P}\Big(\sum_{w \in W_{ab}(I_{ab})} V_w(B) < \delta \log n\Big)\Big)^{K_d}. \qquad (37)
\end{aligned}$$

Since for all $a \neq b$,

$$V(B) \geq \sum_{w \in W_{ab}(I_{ab})} V_w(B) \quad \text{a.s.},$$

and $|\{\{a,b\} \in Z^2 : a \neq b\}| = \binom{k}{2}$, the right-hand side of (37) is bounded by

$$\left(1 + \binom{k}{2}\right)^{K_d} \max_{a \neq b} \mathbb{P}\Big(\sum_{w \in W_{ab}(I_{ab})} V_w(B) < \delta \log n \Big)^{K_d}$$
$$\leq \left(1 + \binom{k}{2}\right)^{K_d} \mathbb{P}\Big(\text{Poisson}(\lambda \pi_{\text{wit}} \chi \ell^d \log n) < \delta \log n \Big)^{K_d}.$$

By Lemma 9 it holds that

$$\mathbb{P}\left(\text{Poisson}(\lambda \pi_{\text{wit}} \chi \ell^d \log n) < \delta \log n\right)^{K_d} \leq n^{-\lambda \pi_{\text{wit}} \ell^d \chi K_d} \Big(\frac{e \lambda \pi_{\text{wit}} \ell^d \chi}{\delta} \Big)^{\delta K_d \log n}.$$

Let $x := ak > 1$ and $y := \delta k$. For $0 < \delta < (ak-1)^2/(8ak^2)$ we have $0 < y < (x-1)^2/(8x)$. Since $y \mapsto y \log(ex/y)$ is increasing for $0 < y < x$,

$$y \log \frac{ex}{y} \leq \frac{(x-1)^2}{8x} \log\Big(\frac{8ex^2}{(x-1)^2} \Big) < \frac{x-1}{2},$$

where the last inequality follows by writing $r = (x-1)/x \in (0,1)$ and using $r \log(8e/r^2) < 4$ (the left-hand side is increasing on $(0,1)$ and is $\log(8e) < 4$ at $r = 1$). Thus $(ea/\delta)^{\delta k} < e^{(ak-1)/2}$. Taking $a := \lambda \pi_{\text{wit}} \ell^d \chi$ and $k := K_d$, for all $0 < \delta < \delta_0$,

$$\mathbb{P}\left(\text{Poisson}(\lambda \pi_{\text{wit}} \chi \ell^d \log n) < \delta \log n\right)^{K_d} \leq n^{-(1 + \lambda \pi_{\text{wit}} \ell^d \chi K_d)/2} \in o(1/n).$$

Thus,

$$\mathbb{P}(Y_\delta \geq K_d) \leq O(n/\log n) \cdot o(1/n) = o(1).$$

□

In the following lemma we consider the subgraph H_{spread} of H, induced by all δ-spread blocks.

Lemma 23. *(cf. Proposition C.9 in Gaudio et al. (2025)) Suppose that $d \geq 2$ and $\lambda \pi_{\text{wit}} \nu_d \ell^d > 1$. Let $V \subset \mathcal{S}_{d,n}$ be a Poisson point process on $\mathcal{S}_{d,n}$ with intensity λ. Then for χ and δ given in* (9) *and* (13)*, respectively, the $(\chi \ell^d \log n, \delta \log n)$-proximity-spread graph H_{spread} of G is connected with probability $1 - o(1)$.*

Combining Lemma 22 and Lemma 23, we obtain that every block in H has a neighboring block in H_{spread}. This property of H_{spread} will be useful in our propagation proof.

Proposition 24. *Suppose that for $d = 1$, $\lambda \pi_{\text{wit}} \ell > 1$, and, for $d \geq 2$, $\lambda \pi_{\text{wit}} \nu_d \ell^d > 1$. Let $V \subset \mathcal{S}_{d,n}$ be a Poisson point process on $\mathcal{S}_{d,n}$ with intensity λ. For δ given by* (13)*, and χ satisfying* (8) *if $d = 1$ and* (9) *if $d \geq 2$, H_{spread} is connected with high probability. Moreover, with high probability, every block has at least one neighbor in H_{spread}.*

Proof. By Lemma 22, with high probability there is no cluster of at least K_d non-δ-spread blocks. On this event, if $d = 1$, disconnectedness of H_{spread} would imply the existence of K_1 consecutive non-δ-spread blocks, a contradiction. Hence H_{spread} is connected when $d = 1$. For $d \geq 2$, connectivity follows from Lemma 23.

It remains to show that every block has a neighbor in H_{spread}. If B is δ-spread, then, since H_{spread} is connected, B has a neighbor in H_{spread} (for sufficiently large n, the graph contains more than one vertex).

Now suppose that B is not δ-spread. If no block ℓ-close to B is δ-spread, then B together with all blocks ℓ-close to B forms a cluster of non-δ-spread blocks of cardinality at least K_d. This contradicts the event from Lemma 22. Hence B has an ℓ-close δ-spread block, which is a neighbor of B in H_{spread}. □

H.2 Almost exact recovery in Phase II

Let $\omega^* \in \Omega_{\pi,P}$ be the permissible relabeling corresponding to the estimated labels of the initial block, that is,

$$\hat{\sigma}(v) = \omega^* \circ \sigma^*(v) \qquad \text{for all } v \in V_{\text{init}}.$$

Due to Proposition 13, ω^* exists with high probability. Since ω^* is permissible, the model, the propagation rule, and the affinity information are invariant under composition with $(\omega^*)^{-1}$. Hence, throughout the proof of Phase II, we may without loss of generality assume that

$$\omega^* = \mathrm{id}.$$

The following lemma bounds the probability of mislabeling a vertex $v \in V$ in terms of the number of errors among its labeled neighbors and the affinity information available for v.

Lemma 25 (Propagation error bound)**.** *Let $G \sim \mathrm{GHCM}(\lambda, n, r, \pi, P(\cdot), d)$, $u \in V$, $S \subset V \setminus \{u\}$ and $\hat{\sigma}_S : S \to Z$ be a labeling on S. Let $\hat{\sigma}(u)$ be the output of Propagate (Algorithm 2) applied on input $(G, S, \{u\}, \hat{\sigma}_S)$, assume that Assumption 3 holds with some constant $\rho > 0$. Let $\mathcal{F}$ be a σ-field with respect to which $(V, \sigma^*, \widehat{\sigma}_S)$ is measurable and such that, conditionally on (V, σ^*), the random variables*

$$\{X_{uv} : v \in S \cap N(u)\}$$

are independent of $\mathcal{F}$. Then,

$$\mathbb{P}(\hat{\sigma}(u) \neq \omega^* \circ \sigma^*(u) \mid \mathcal{F}) \leq (k-1) e^{\rho m_{\hat{\sigma}}(S) - \mathcal{I}_S(u)},$$

where $m_{\hat{\sigma}}(S) := |\{v \in S \cap N(u) : \ \omega^* \circ \sigma^*(v) \neq \hat{\sigma}_S(v)\}|$ is the number of incorrectly labeled nodes in S and $\mathcal{I}_S(u)$ is given at (2)*.*

For Algorithms A and B, the condition on $\mathcal{F}$ is satisfied with $\mathcal{F} = \sigma(V, \sigma^*, \hat{\sigma}_S)$, since an edge incident to an unlabeled vertex is not used before that vertex is labeled.

Proof. Let $a := \omega^* \circ \sigma^*(u)$. Note that, on the event $\{a \neq \hat{\sigma}(u)\}$, the likelihood of u having label $b \in Z \setminus \{a\}$ exceeds the likelihood of u having label a, i.e.

$$\Lambda_b := \sum_{v \in S \cap N(u)} \log \frac{\bar{p}_{b\hat{\sigma}(v)}(X_{uv}; \|u - v\|)}{\bar{p}_{a\hat{\sigma}(v)}(X_{uv}; \|u - v\|)} \geq 0,$$

where we write $\hat{\sigma}(v) := \hat{\sigma}_S(v)$ for $v \in S$. Then, by Markov's inequality, we have for all $t \in (0, 1)$,

$$\begin{aligned}
\mathbb{P}(\Lambda_b \geq 0 \mid \mathcal{F}) &= \mathbb{P}(e^{t\Lambda_b} \geq 1 \mid \mathcal{F}) \\
&\leq \mathbb{E}(e^{t\Lambda_b} \mid \mathcal{F}) \\
&= \prod_{v \in S \cap N(u)} \int_{\mathbb{X}} \Big(\frac{\bar{p}_{b\hat{\sigma}(v)}(x; \|u - v\|)}{\bar{p}_{a\hat{\sigma}(v)}(x; \|u - v\|)} \Big)^t \bar{P}_{a, \omega^* \circ \sigma^*(v)}(\|u - v\|)(\mathrm{d}x).
\end{aligned}$$

Fix $v \in S \cap N(u)$ and write $w := \omega^* \circ \sigma^*(v)$ for its true label and $w' := \hat{\sigma}(v)$ for its assigned one, so that the v-th factor above is

$$J_v := \int_{\mathbb{X}} \Big(\frac{\bar{p}_{bw'}(x; \|u - v\|)}{\bar{p}_{aw'}(x; \|u - v\|)} \Big)^t \bar{p}_{aw}(x; \|u - v\|) \, \mu(\mathrm{d}x).$$

If $w = w'$, then $J_v = \phi_t\big(\bar{P}_{bw'}(\|u - v\|), \bar{P}_{aw'}(\|u - v\|)\big)$ by (1). If $w \neq w'$, Assumption 3 gives $\bar{p}_{aw}(x; y) \leq e^{\rho} \, \bar{p}_{aw'}(x; y)$ for all $x \in \mathbb{X}$ and all $y \in [0, r]$, whence

$$J_v \leq e^{\rho} \, \phi_t\big(\bar{P}_{bw'}(\|u - v\|), \bar{P}_{aw'}(\|u - v\|)\big).$$

There are at most $m_{\hat{\sigma}}(S)$ indices of the second kind. Taking $t = 1/2$, for which $\phi_{1/2}$ is symmetric in its arguments, gives

$$\mathbb{P}(\Lambda_b \geq 0 \mid \mathcal{F}) \leq e^{\rho m_{\hat{\sigma}}(S)} \prod_{v \in S \cap N(u)} \phi_{1/2}\Big(\bar{P}_{b\hat{\sigma}(v)}(\|u - v\|), \bar{P}_{a\hat{\sigma}(v)}(\|u - v\|)\Big).$$

Using Definition 2 and a union bound over all $b \in Z \setminus \{a\}$ gives the asserted bound. □

Our next goal is to establish a uniform upper bound on the number of mislabeled nodes that are not labeled $*$ in the neighborhood of every vertex. We will later use this bound to obtain a logarithmic upper bound on the total number of mislabelings in every neighborhood.

Proposition 26 (Uniform error bound)**.** *Assume that Assumptions 1 and 3 hold with some constant $\rho > 0$, and that Algorithm A or Algorithm B is used. There exists $m \in \mathbb{N}$ such that*

$$\mathbb{P}\Big(\bigcap_{w \in V} \{ |\{\nu \in V \cap N(w) : \widehat{\sigma}(\nu) \notin \{\omega^* \circ \sigma^*(\nu), *\}\}| \le m \} \Big) = 1 - o(1).$$

The main idea is an induction on the Phase II history. Once the source of the next vertex contains only a bounded number of errors, the geometric argument below guarantees at least a constant multiple of $\log n$ affinity information for both propagation mechanisms. Lemma 25 then makes the next error polynomially unlikely.

Definition 27 (Proximity chain)**.** A sequence of blocks $\mathcal{P} = (B_1, \dots, B_k)$ is called a proximity chain if, for every $i < k$,

$$\sup_{x \in B_i, z \in B_{i+1}} \|x - z\| \le \ell (\log n)^{1/d}.$$

The next lemma connects proximity chains to the intervals used to identify witness communities. For $v \in \mathcal{S}_{d,n}$ and $I \subseteq [0, r]$ recall the definition

$$\bar{\mathcal{A}}_I(v) := \Big\{ u \in \mathcal{S}_{d,n} : \frac{\|v - u\|}{(\log n)^{1/d}} \in I \Big\}.$$

Lemma 28 (Past witness block)**.** *Let $v \in B$ and let $\mathcal{P} = (B_1, \dots, B_i = B)$ be a proximity chain such that*

$$\sup_{x \in \bigcup_{j<i} B_j} \|x - v\| \ge r (\log n)^{1/d}.$$

Let $I = [\alpha, \beta] \subseteq [0, r]$ satisfy $\beta - \alpha \ge \ell$. Then there exists $j < i$ such that $B_j \subseteq \bar{\mathcal{A}}_I(v)$.

Proof. If $\alpha = 0$, take $j = i - 1$. Since $v \in B_i$ and $\mathcal{P}$ is a proximity chain, for every $x \in B_{i-1}$,

$$\|v - x\| \le \ell (\log n)^{1/d} \le \beta (\log n)^{1/d},$$

and hence $B_{i-1} \subseteq \bar{\mathcal{A}}_I(v)$.

Assume now that $\alpha > 0$, and let

$$j^* = \max \left\{ j < i : \inf_{x \in B_j} \|v - x\| \ge \alpha (\log n)^{1/d} \right\}.$$

The set is nonempty. Indeed, by assumption there exist $j_0 < i$ and $y \in B_{j_0}$ such that

$$\|v - y\| \ge r (\log n)^{1/d}.$$

For x, y in the same block we have

$$\|x - y\| \le \sqrt{d} \ell \chi^{1/d} (\log n)^{1/d} < \ell (\log n)^{1/d}.$$

Hence, for every $x \in B_{j_0}$, since $r - \ell \ge \alpha$,

$$\|v - x\| \ge \|v - y\| - \|x - y\| \ge (r - \ell)(\log n)^{1/d} \ge \alpha (\log n)^{1/d}.$$

If $j^* < i - 1$, by maximality there exists $z \in B_{j^*+1}$ such that $\|v - z\| < \alpha (\log n)^{1/d}$. If $j^* = i - 1$, take $z = v \in B_i$, for which the same inequality holds.

For every $x \in B_{j^*}$, since $\mathcal{P}$ is a proximity chain,

$$\|v - x\| \le \|v - z\| + \|z - x\| \le (\alpha + \ell)(\log n)^{1/d} \le \beta (\log n)^{1/d}.$$

By the definition of j^*, $\|v - x\| \ge \alpha (\log n)^{1/d}$. Hence $B_{j^*} \subseteq \bar{\mathcal{A}}_I(v)$. Since $j^* < i$, this block belongs to the past of v. □

Lemma 29 (Initial spread core)**.** *Let*

$$\mathcal{P}_{\text{init}} := \{B \in \mathcal{P}_\chi : B \subseteq B_{\text{init}}\}.$$

With high probability, every block in $\mathcal{P}_{\text{init}}$ is δ-spread. Moreover, the subgraph of H_{spread} induced by $\mathcal{P}_{\text{init}}$ is connected.

Proof. Write $L_n = (\log n)^{1/d}$ and let $h_n = \ell\chi^{1/d}L_n$ be the block side length. Since $B_{\text{init}} = [-rL_n, rL_n]^d$, the number of blocks fully contained in B_{init} is bounded uniformly in n.

For any fixed block B, the Poisson lower-tail estimates used in the proof of Lemma 22 imply

$$\mathbb{P}(B \text{ is not } \delta\text{-spread}) = o(1).$$

A union bound over the $O(1)$ blocks of $\mathcal{P}_{\text{init}}$ therefore shows that all of them are δ-spread with probability $1 - o(1)$.

The blocks in $\mathcal{P}_{\text{init}}$ form a rectangular grid. If two such blocks share a $(d-1)$-dimensional face, then

$$\sup_{x\in B,\, y\in B'} \|x - y\| \le \sqrt{d+3}\, h_n < \ell L_n$$

by (10). Hence face-adjacent blocks are ℓ-close, and the induced subgraph is connected. □

Lemma 30 (Spread proximity chain)**.** *Assume that Algorithm A is used. Then, with high probability, at every step of Phase II before all vertices have been labeled, there exist a not fully labeled block B and fully labeled, δ-spread blocks $B_1, \dots, B_i$ such that $(B_1, \dots, B_i, B)$ is a proximity chain and, for every $v \in B$,*

$$\sup_{x\in\bigcup_{j=1}^{i} B_j} \|x - v\| \ge r(\log n)^{1/d}. \tag{38}$$

Proof. Write

$$L_n := (\log n)^{1/d}, \qquad h_n := \ell\chi^{1/d}L_n.$$

We work on the intersection of the high-probability events from Lemma 29 and Proposition 24. In particular, H_{spread} is connected, every block has an ℓ-close δ-spread block, and the initial spread blocks form a connected set.

Fix an arbitrary step of Phase II before termination. Let $\mathcal{C}$ be the connected component, in the graph induced by the currently fully labeled δ-spread blocks, that contains $\mathcal{P}_{\text{init}}$. These initial blocks are fully labeled because Phase I labels every vertex of B_{init}.

Since Phase II has not terminated, choose a currently unlabeled vertex z and let D be its block. By Proposition 24, there is a δ-spread block D' that is ℓ-close to D.

If $D' \in \mathcal{C}$, set $B := D$. Then B is not fully labeled and is ℓ-close to $\mathcal{C}$.

Suppose instead that $D' \notin \mathcal{C}$. Since H_{spread} is connected, take a path of δ-spread blocks

$$C_0, C_1, \dots, C_s = D'$$

with $C_0 \in \mathcal{P}_{\text{init}}$, and let

$$j_* := \min\{j : C_j \notin \mathcal{C}\}.$$

Then $j_* \ge 1$, $C_{j_*-1} \in \mathcal{C}$, and C_{j_*} is ℓ-close to C_{j_*-1}. Set $B := C_{j_*}$. The block B cannot be fully labeled: if it were, then, being δ-spread and adjacent to $\mathcal{C}$, it would itself belong to $\mathcal{C}$, contradicting the definition of j_*. Thus in either case we obtain a not fully labeled block B that is ℓ-close to $\mathcal{C}$.

Choose a currently unlabeled vertex $u \in B$. Since every vertex of B_{init} was labeled in Phase I, $u \notin B_{\text{init}}$. Hence, for some coordinate $q \in \{1, \dots, d\}$ and some $\varepsilon \in \{-1, 1\}$,

$$\varepsilon u_q > rL_n.$$

For every $v \in B$ we have $|u_q - v_q| \le h_n$, and therefore

$$\varepsilon v_q > rL_n - h_n. \tag{39}$$

Choose a block $B_0 \in \mathcal{P}_{\mathrm{init}}$ adjacent to the opposite q-face of B_{init}. Because the mesh size is h_n, we may choose $x_0 \in B_0$ such that

$$-\varepsilon(x_0)_q \geq rL_n - h_n.$$

For all sufficiently large n, the sets under consideration lie far from the torus identification, so the ordinary Euclidean coordinate bound agrees with the toroidal metric. Combining the last two displays yields, for every $v \in B$,

$$\|x_0 - v\| \geq |(x_0)_q - v_q| > 2rL_n - 2h_n \geq rL_n,$$

where the last inequality follows from (10). Hence

$$\sup_{x \in B_0} \|x - v\| \geq r(\log n)^{1/d} \qquad \text{for every } v \in B. \tag{40}$$

Both B_0 and the spread block of $\mathcal{C}$ adjacent to B belong to $\mathcal{C}$. Since $\mathcal{C}$ is connected, choose a path

$$B_1 = B_0, B_2, \ldots, B_i$$

inside $\mathcal{C}$ whose final block B_i is ℓ-close to B. Every B_j is fully labeled and δ-spread, and $(B_1, \ldots, B_i, B)$ is a proximity chain. Finally, (40) implies (38). □

We introduce a partial ordering on V and write $u < v$ for vertices $u, v \in V$ if u is labeled before v by the estimator $\widehat{\sigma}$. Let $\widehat{V}_v := \{u \in V : u < v\}$ be the set of nodes labeled before v.

For a vertex v labeled during Phase II, let $S(v)$ denote the source actually used to label v. Thus, for Algorithm A,

$$S(v) := \widehat{V}_v \cap N(v),$$

whereas for Algorithm B,

$$S(v) := N(v) \cap \bigcup_{j \in \widehat{V}^{\dagger}_{\mathrm{cert}}(v)} (V \cap B_j),$$

where $\widehat{V}^{\dagger}_{\mathrm{cert}}(v)$ denotes the collection of certified blocks available immediately before the block containing v is processed.

Lemma 31 (Enough affinity information)**.** *There exists $c_2 > 0$ such that, for every fixed $m \in \mathbb{N}$, the following holds. Assume that Algorithm A or Algorithm B is used. For $v \in V$, let*

$$\mathcal{G}_v := \bigcap_{w \in V} \left\{ \left| \left\{ \nu \in \widehat{V}_v \cap N(w) : \widehat{\sigma}(\nu) \notin \{\omega^* \circ \sigma^*(\nu), *\} \right\} \right| \leq m \right\}.$$

Then

$$\mathbb{P}\left(\bigcup_{v \in V \setminus V_{\mathrm{init}}} \left\{ \mathcal{G}_v \cap \left\{ \mathcal{I}_{S(v)}(v) \leq c_2 \log n \right\} \right\} \right) = o(1).$$

Proof. Suppose that a vertex v is labeled during Phase II and that

$$\mathcal{G}_v \qquad \text{and} \qquad \mathcal{I}_{S(v)}(v) \leq c_2 \log n$$

hold. We distinguish the two algorithms.

Algorithm A. Work on the high-probability event of Lemma 30. Let B and fully labeled δ-spread blocks $B_1, \ldots, B_i$ be supplied by that lemma, and choose any currently unlabeled vertex $u \in B$. Since Algorithm A chooses v to maximize the affinity information among all currently unlabeled vertices,

$$\mathcal{I}_{\widehat{V}_v \cap N(u)}(u) \leq \mathcal{I}_{\widehat{V}_v \cap N(v)}(v) = \mathcal{I}_{S(v)}(v) \leq c_2 \log n.$$

Define

$$T(u) := \widehat{V}_v \cap N(u) \cap \bigcup_{j=1}^{i} B_j.$$

Since $T(u) \subseteq \widehat{V}_v \cap N(u)$ and each summand in Definition 2 is nonnegative,

$$\mathcal{I}_{T(u)}(u) \le \mathcal{I}_{\widehat{V}_v \cap N(u)}(u) \le c_2 \log n.$$

For each $j \le i$, the block B_j is δ-spread and hence contains at least one vertex. Choose $y_j \in V \cap B_j$. By (10), the diameter of a block is smaller than the visibility radius, so every other vertex of B_j lies in $N(y_j)$. On $\mathcal{G}_v$, at most m previously labeled vertices of $N(y_j)$ are incorrect; allowing also for y_j itself, B_j contains at most $m+1$ incorrectly labeled vertices. Therefore, for every $a \neq b$ and all sufficiently large n,

$$\sum_{w \in W_{ab}(I_{ab})} |\{x \in V \cap B_j : \widehat{\sigma}(x) = w\}| \ge \frac{3}{4}\delta\pi_{\mathrm{wit}} \log n - (m+1) \tag{41}$$

$$\ge \frac{1}{2}\delta\pi_{\mathrm{wit}} \log n. \tag{42}$$

Moreover, (38) of Lemma 30 gives

$$\sup_{x \in \bigcup_{j=1}^{i} B_j} \|x - u\| \ge r(\log n)^{1/d},$$

since $u \in B$. The two branches have therefore produced the same three ingredients, and the remainder of the argument is common to both.

Algorithm B. Set $u := v$, and let B be the block containing v. By construction of Algorithm B, B is ℓ-close to a previously labeled certified block. Iterating the predecessor relation, and extending the chain inside the initially labeled certified core if necessary, we obtain fully labeled certified blocks $B_1, \dots, B_i$ such that $(B_1, \dots, B_i, B)$ is a proximity chain and

$$\sup_{x \in \bigcup_{j=1}^{i} B_j} \|x - u\| \ge r(\log n)^{1/d}.$$

Define

$$T(u) := N(u) \cap \bigcup_{j=1}^{i} (V \cap B_j).$$

All blocks $B_1, \dots, B_i$ are certified and available before the block containing u is processed. Consequently $T(u) \subseteq S(u)$ and

$$\mathcal{I}_{T(u)}(u) \le \mathcal{I}_{S(u)}(u) = \mathcal{I}_{S(v)}(v) \le c_2 \log n.$$

Furthermore, by certification, for every $a \neq b$ and every $j \le i$,

$$\sum_{w \in W_{ab}(I_{ab})} |\{x \in V \cap B_j : \widehat{\sigma}(x) = w\}| \ge \frac{1}{2}\delta\pi_{\mathrm{wit}} \log n. \tag{43}$$

Thus we have constructed the current vertex u, previously labeled blocks $B_1, \dots, B_i$, and a set $T(u)$ such that

$$\mathcal{I}_{T(u)}(u) \le c_2 \log n, \qquad \sup_{x \in \bigcup_{j=1}^{i} B_j} \|x - u\| \ge r(\log n)^{1/d},$$

and, for every $a \neq b$ and every $j \le i$,

$$\sum_{w \in W_{ab}(I_{ab})} |\{x \in V \cap B_j : \widehat{\sigma}(x) = w\}| \ge \frac{1}{2}\delta\pi_{\mathrm{wit}} \log n. \tag{44}$$

By Lemma 28, for every $a \neq b$ there exists $j(a,b) \le i$ such that $B_{j(a,b)} \subseteq \bar{\mathcal{A}}_{I_{ab}}(u)$. Fix $t = 1/2$ and $\Phi \in (0,1)$. Then

$$\mathcal{I}_{T(u)}(u) \ge \min_{a \neq b} \sum_{x \in V \cap B_{j(a,b)}} -\log \phi_{1/2}\left(\bar{P}_{a\widehat{\sigma}(x)}(\|u-x\|), \bar{P}_{b\widehat{\sigma}(x)}(\|u-x\|)\right).$$

Hence, if $\mathcal{I}_{T(u)}(u) \le c_2 \log n$, then, for some $a \neq b$, at most

$$\frac{c_2}{\log(1/\Phi)} \log n$$

of the vertices counted in (44) can satisfy

$$\phi_{1/2}\left(\bar{P}_{a\widehat{\sigma}(x)}(\|u-x\|), \bar{P}_{b\widehat{\sigma}(x)}(\|u-x\|)\right) \le \Phi.$$

Consequently, at least

$$\left(\frac{1}{2}\delta\pi_{\text{wit}} - \frac{c_2}{\log(1/\Phi)}\right) \log n$$

vertices in $B_{j(a,b)}$ satisfy

$$\phi_{1/2}\left(\bar{P}_{a\widehat{\sigma}(x)}(\|u-x\|), \bar{P}_{b\widehat{\sigma}(x)}(\|u-x\|)\right) > \Phi.$$

Define

$$S(u,\Phi) := \bigcup_{a\neq b} \bigcup_{w \in W_{ab}(I_{ab})} \left\{x \in \bar{\mathcal{A}}_{I_{ab}}(u) : \phi_{1/2}\left(\bar{P}_{aw}(\|u-x\|), \bar{P}_{bw}(\|u-x\|)\right) > \Phi\right\}.$$

It follows that

$$V(S(u,\Phi)) \ge \left(\frac{1}{2}\delta\pi_{\text{wit}} - \frac{c_2}{\log(1/\Phi)}\right) \log n.$$

Choosing

$$c_2 < \frac{1}{4}\delta\pi_{\text{wit}} \log(1/\Phi)$$

gives

$$V(S(u,\Phi)) \ge \frac{1}{4}\delta\pi_{\text{wit}} \log n.$$

By Lemma 11, for every $\epsilon > 0$ there exists $\Phi < 1$ such that, uniformly in u,

$$|S(u,\Phi)| \le \epsilon \log n.$$

Thus, by the Mecke equation,

$$\mathbb{P}\left(\bigcup_{u\in V\setminus V_{\text{init}}} \left\{V(S(u,\Phi)) \ge \frac{1}{4}\delta\pi_{\text{wit}} \log n\right\}\right) \le \lambda \int_{\mathcal{S}_{d,n}\setminus B_{\text{init}}} \mathbb{P}\left(V(S(u,\Phi)) \ge \frac{1}{4}\delta\pi_{\text{wit}} \log n\right) \mathrm{d}u.$$

Since $V(S(u,\Phi))$ is Poisson with mean $\lambda|S(u,\Phi)| \le \lambda\epsilon \log n$, the last expression is bounded by

$$\lambda|\mathcal{S}_{d,n} \setminus B_{\text{init}}|\, \mathbb{P}\left(\text{Poisson}(\lambda\epsilon \log n) \ge \frac{1}{4}\delta\pi_{\text{wit}} \log n\right),$$

which is $o(1)$ for $\epsilon > 0$ sufficiently small. □

Proof of Proposition 26. Let $c_0 := c_2$, where $c_2 > 0$ is the constant from Lemma 31, and choose $m \in \mathbb{N}$ sufficiently large that $c_0(m+1) > 1$. Let

$$\mathcal{G}_v := \bigcap_{w\in V} \{\left|\left\{u \in \widehat{V}_v \cap N(w) : \widehat{\sigma}(u) \notin \{\omega^* \circ \sigma^*(u), *\}\right\}\right| \le m\}.$$

We define the auxiliary events

$$\begin{aligned}
\mathcal{E}_0 &:= \bigcap_{u\in V_{\text{init}}} \{\widehat{\sigma}(u) = \omega^* \circ \sigma^*(u)\}, \\
\mathcal{H} &:= \bigcap_{v\in V\setminus V_{\text{init}}} \left(\mathcal{G}_v^c \cup \left\{\mathcal{I}_{S(v)}(v) \ge c_0 \log n\right\}\right), \\
\mathcal{A} &:= \{\text{all blocks contain at most } \Delta \log n \text{ vertices}\}.
\end{aligned}$$

Then $\mathbb{P}(\mathcal{E}_0 \cap \mathcal{H} \cap \mathcal{A}) = 1 - o(1)$. Let $v_1, v_2, \dots$ be the Phase II labeling order and let $\mathcal{F}_{t-1}$ contain V, σ^*, and the complete history up to just before v_t is labeled. The selected vertex v_t and its current source are $\mathcal{F}_{t-1}$-measurable, while the edge variables incident to v_t that are used by `Propagate` have not previously been revealed. On the event that every neighborhood contains at most m earlier errors, Lemma 31 supplies the required information lower bound for either propagation mechanism. Lemma 25 therefore gives

$$\mathbb{P}\left(\widehat{\sigma}(v_t) \notin \{\omega^* \circ \sigma^*(v_t), *\} \,|\, \mathcal{F}_{t-1}\right) \le p_n, \tag{45}$$

where $p_n = C_0 n^{-c_0}$ for constants $C_0 < \infty$ and $c_0 > 0$.

We use the following elementary predictable-thinning fact. If $Z_t \in \{0,1\}$ is adapted and $\mathbb{P}(Z_t = 1 \mid \mathcal{F}_{t-1}) \le p$, while $A_t \in \{0,1\}$ is $\mathcal{F}_{t-1}$-measurable, then, on any event on which $\sum_t A_t \le L$,

$$\mathbb{P}\left(\sum_t A_t Z_t \ge q\right) \le \mathbb{P}(\mathrm{Bin}(L,p) \ge q).$$

This follows by exposing only the predictable times for which $A_t = 1$ and iterating the tower property.

For a fixed w, take $A_t = \mathbf{1}\{v_t \in N(w)\}$ and let Z_t indicate a new non-$*$ propagation error while the local-error bound and the relevant information lower bound hold. Since v_t is selected before its incident edge variables are revealed, A_t is $\mathcal{F}_{t-1}$-measurable. On $\mathcal{A}$, at most $C \log n$ vertices lie in $N(w)$. Hence

$$\mathbb{P}\left(\sum_t A_t Z_t > m \,\middle|\, \mathcal{A}\right) \le \binom{C\log n}{m+1} p_n^{m+1}.$$

By increasing m if necessary, we may fix $\varepsilon > 0$ such that $c_0(m+1) > 1 + \varepsilon$. Then, uniformly in w,

$$\mathbb{P}\left(\mathcal{A} \cap \left\{\sum_t A_t Z_t > m\right\}\right) = o(n^{-1}). \tag{46}$$

It remains to connect the predictably thinned error variables to the total number of propagation errors in every neighborhood.

Suppose that the conclusion of Proposition 26 fails. Let

$$v_1, v_2, \dots$$

denote the vertices in their Phase II labeling order, and define

$$t^* := \min\left\{t : \exists\, w \in V : |\{v_j : j \le t,\, v_j \in N(w),\, \widehat{\sigma}(v_j) \notin \{\omega^* \circ \sigma^*(v_j), *\}\}| \ge m+1\right\}.$$

Set

$$v^* := v_{t^*}.$$

By the minimality of t^*, before each vertex v_j with $j \le t^*$ is labeled, every neighborhood contains at most m incorrectly labeled vertices. Hence

$$\mathcal{G}_{v_j} \qquad \text{holds for every } j \le t^*.$$

Let $w \in V$ witness the definition of t^*. By minimality of t^*, every one of the first $m+1$ erroneous vertices in $N(w)$ was created while the local-error hypothesis still held. The appropriate information lower bound therefore held at each of those labeling times, so the corresponding variables $A_t Z_t$ all equal one. Consequently

$$\sum_t A_t Z_t > m.$$

It follows that

$$\mathcal{E}_0 \cap \mathcal{H} \cap \mathcal{A} \cap \{\exists\, w \in V : |\{u \in V \cap N(w) : \widehat{\sigma}(u) \notin \{\omega^* \circ \sigma^*(u), *\}\}| > m\}$$

is contained in

$$\bigcup_{w \in V} \left\{\sum_t \mathbf{1}\{v_t \in N(w)\} Z_t > m\right\}.$$

Hence

$$\mathbb{P}\left(\exists\, w \in V : |\{u \in V \cap N(w) : \widehat{\sigma}(u) \notin \{\omega^* \circ \sigma^*(u), *\}\}| > m\right)$$
$$\leq \mathbb{P}\big((\mathcal{E}_0 \cap \mathcal{H} \cap \mathcal{A})^c\big) + \mathbb{P}\left(\mathcal{A} \cap \bigcup_{w \in V} \left\{\sum_t \mathbf{1}\{v_t \in N(w)\} Z_t > m\right\}\right).$$

The first term is $o(1)$. Moreover, since $|V| \leq \Delta n$ on $\mathcal{A}$, (46) and the Mecke equation give

$$\mathbb{P}\left(\mathcal{A} \cap \bigcup_{w \in V} \left\{\sum_t \mathbf{1}\{v_t \in N(w)\} Z_t > m\right\}\right) = o(1).$$

Therefore

$$\mathbb{P}\left(\bigcap_{w \in V} \{|\{u \in V \cap N(w) : \widehat{\sigma}(u) \notin \{\omega^* \circ \sigma^*(u), *\}\}| \leq m\}\right) = 1 - o(1),$$

which proves the proposition. □

Finally, we conclude that after Phase II, there are at most $\beta \log n$ incorrectly labeled vertices in the neighborhood of any vertex. Note that, in contrast to the previous statements, the following corollary also allows us to control the number of vertices labeled $*$.

Corollary 32. *Assume that Algorithm A or Algorithm B is used under the hypotheses of Proposition 26. For every $\beta > 0$, and for Algorithm B with $\delta > 0$ chosen sufficiently small,*

$$\mathbb{P}\Big(\bigcap_{u \in V} \{|\{v \in V \cap N(u) :\ \hat{\sigma}(v) \neq \omega^* \circ \sigma^*(v)\}| \leq \beta \log n\}\Big) = 1 - o(1).$$

Proof. For Algorithm A, Phase II does not output $*$ with high probability, so Proposition 26 gives at most m errors in every neighborhood. Hence, for any fixed $\beta > 0$, $m \leq \beta \log n$ for all sufficiently large n.

For Algorithm B, the neighborhood of every $u \in \mathcal{S}_{d,n}$ intersects at most K_r blocks. Every δ-occupied block is eventually processed with high probability, while vertices in non-occupied blocks may retain label $*$. Thus at most $K_r \delta \log n$ such vertices occur in any neighborhood. Together with Proposition 26,

$$|\{v \in V \cap N(u) : \widehat{\sigma}(v) \neq \omega^* \circ \sigma^*(v)\}| \leq K_r \delta \log n + m.$$

Choose δ so that $K_r \delta \leq \beta/2$ and then take n large enough that $m \leq \beta \log n/2$. This proves the local statement. No global almost-exact-recovery count is needed for refinement. □

Proposition 33 (Propagation runtime)**.** *Fix $k = |Z|$ and assume each evaluation of $\phi_{1/2}$ and of a log-density $\log \bar{p}_{ab}(x; y)$ costs $O(1)$. Then, with high probability, Phase II of Algorithm A performs $O(n \log^2 n)$ operations, given the visible pairs and their weights.*

Proof. For $u \in \mathcal{S}_{d,n}$, $|V \cap N(u)|$ is Poisson with mean $\lambda \nu_d r^d \log n$. By Lemma 9 and the Mecke equation (Proposition 8), exactly as in Corollary 10, there is $\Delta > 0$ with

$$\mathbb{P}\Big(\bigcap_{u \in V} \{|V \cap N(u)| \leq \Delta \log n\}\Big) = 1 - o(1).$$

Work on this event. Summing over $u \in V$ and using $|V| = O(n)$ with high probability, the number of visible pairs is $O(n \log n)$.

Maintain the unlabeled vertices in a max-heap keyed by $\mathcal{I}_{\widehat{V} \cap N(v)}(v)$, storing with each v the $\binom{k}{2}$ partial sums

$$s_{ab}(v) := \sum_{w \in \widehat{V} \cap N(v)} -\log \phi_{1/2}\big(\bar{P}_{a\widehat{\sigma}(w)}(\|v - w\|), \bar{P}_{b\widehat{\sigma}(w)}(\|v - w\|)\big), \qquad a \neq b,$$

so that the key is $\min_{a\neq b} s_{ab}(v)$. Initialization costs $O(n)$ heap insertions.

Each iteration extracts the maximum, at cost $O(\log n)$, and labels the extracted vertex u by Algorithm 2, which sums $|\widehat{V} \cap N(u)| \leq \Delta \log n$ log-densities for each of the k candidate labels, hence $O(k \log n)$ operations. Because the exponent $t = 1/2$ is fixed, the newly assigned label $\widehat{\sigma}(u)$ enters each score of each unlabeled neighbor additively: for every $v \in N(u)$ still unlabeled, each $s_{ab}(v)$ is incremented by one evaluation of $\phi_{1/2}$, the key is recomputed as a minimum over $\binom{k}{2}$ values, and the heap is updated in $O(\log n)$. There are at most $\Delta \log n$ such v, so the iteration costs $O\big(k^2 \Delta \log n + \Delta \log^2 n\big)$.

Phase II performs at most $|V| = O(n)$ iterations, giving $O\big(n(k^2 \log n + \log^2 n)\big) = O(n \log^2 n)$ for fixed k. Finally, if only the vertex locations are given, bucket them into cubes of side $r(\log n)^{1/d}$: bucketing costs $O(n)$, and each $u \in V$ is compared only with the 3^d cubes adjacent to its own, whose vertex count is $O(\log n)$ in expectation by the same Poisson bound. The visible pairs are therefore enumerated in $\Theta(n \log n)$ expected time, which is optimal since they must all be listed. □

I REFINEMENT

The proof of exact recovery in the Refinement step is analogous to (Gaudio & Jin, 2026, Section C.3). For completeness, we reproduce the argument.

Lemma 34 (Lemma 9 in Gaudio & Jin (2026)). *If $\lambda \nu_d r^d \min_{a\neq b} D_+(P_a \| P_b) > 1$, then for a fixed*

$$0 < \epsilon \leq \frac{1}{2}(\lambda \nu_d r^d \min_{a\neq b} D_+(P_a \| P_b) - 1),$$

we have that

$$\mathbb{P}(\ell_i(v, \sigma^*) - \ell_j(v, \sigma^*) \leq \epsilon \log n \mid \sigma^*(v) = i) = n^{-(1+\Omega(1))}$$

where

$$\ell_i(v, \sigma) = \sum_{\substack{u \in N(v) \\ \sigma(u) \neq *}} \log(\bar{p}_{i,\sigma(u)}(x_{uv}; \|u - v\|)).$$

We finally complete the proof of Theorem 3 following the proof of Theorem 2 in Gaudio & Jin (2026) (see also the proof of Theorem E.3 in Gaudio et al. (2025)).

Completion of the recovery proofs. Recall that ω^* is such that $\hat{\sigma}_0 = \omega^* \circ \sigma^*$, where $\hat{\sigma}_0$ is the initial block labeling. Fix $\beta > 0$ and let the event $\mathcal{E}_1$ be the event that $\hat{\sigma}$ makes at most $\beta \log n$ mistakes in the neighborhood of every vertex $v \in V$. That is,

$$\mathcal{E}_1 = \bigcap_{v \in V} \Big\{ |\{u \in V \cap N(v) : \hat{\sigma}(u) \neq \omega^* \circ \sigma^*(u)\}| \leq \beta \log n \Big\}.$$

By Corollary 32, $\mathbb{P}(\mathcal{E}_1) = 1 - o(1)$.

We will provide a uniform upper bound on the failure probability of the MAP estimator on a given vertex $v \in V$, over all labelings σ which are "close" to the true labels σ^*. Thus, fix $v \in V$ and let $W(v, \beta)$ be the set of labelings σ which differ from $\omega^* \circ \sigma^*$ by at most $\beta \log n$ vertices in the neighborhood of v, meaning that

$$W(v; \beta) := \{\sigma : |\{u \in V \cap N(v) : \sigma(u) \neq \omega^* \circ \sigma^*(u)\}| \leq \beta \log n\}.$$

Now, we define the event

$$\mathcal{E}_v = \bigcup_{i \in Z} \left[\{\omega^* \circ \sigma^*(v) = i\} \cap \left(\bigcup_{\sigma \in W(v;\beta)} \bigcup_{j \neq i} \{\ell_i(v, \sigma) \leq \ell_j(v, \sigma)\} \right) \right],$$

which is the event that there exists a labeling $\sigma \in W(v; \beta)$ such that the MAP estimator with respect to σ fails on v. Note that the probability of exact recovery is bounded by

$$\mathbb{P}(\tilde{\sigma} = \omega^* \circ \sigma^*) \geq 1 - \mathbb{P}(\mathcal{E}_1^c) - \mathbb{P}\Big(\bigcup_{v \in V} \mathcal{E}_v \Big).$$

By the Mecke equation (Proposition 8),

$$\mathbb{P}\Big(\bigcup_{v\in V}\mathcal{E}_v\Big) \le \lambda \int_{\mathcal{S}_{d,n}} \mathbb{P}(\mathcal{E}_v)\,\mathrm{d}v.$$

Next we bound $\mathbb{P}(\mathcal{E}_v)$. By Assumption 3, for every $\sigma \in W(v;\beta)$ and every $i \neq j$,

$$\begin{aligned}&\big|\big(\ell_i(v,\sigma) - \ell_j(v,\sigma)\big) - \big(\ell_i(v,\omega^*\circ\sigma^*) - \ell_j(v,\omega^*\circ\sigma^*)\big)\big| \\ &\qquad \le 2\rho\,|\{u \in V \cap N(v) : \sigma(u) \neq \omega^*\circ\sigma^*(u)\}| \le 2\rho\beta\log n,\end{aligned}$$

where we have used $\omega^* = \mathrm{id}$ without loss of generality. Thus, the event $\bigcup_{\sigma\in W(v;\beta)}\{\ell_i(v,\sigma) \le \ell_j(v,\sigma)\}$ occurring within $\mathcal{E}_v$ implies

$$\{\ell_i(v,\omega^*\circ\sigma^*) \le \ell_j(v,\omega^*\circ\sigma^*) + 2\rho\beta\log n\}. \tag{47}$$

Note that $\beta > 0$ can be arbitrarily small. Conditioned on $\sigma^*(v) = i$, Lemma 34 implies that (47) occurs with probability $n^{-(1+\Omega(1))}$ when $\beta > 0$ is small enough, thus establishing that $\mathcal{E}_v$ holds with probability $n^{-(1+\Omega(1))}$ also. Since $|\mathcal{S}_{d,n}| \in \Theta(n)$, this finishes the proof. □