WEAK SPARSE RELATIONS ADD UP: SHARP LIMITS FOR JOINT COMMUNITY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study community detection in the binary symmetric joint stochastic block model (JSBM), which couples a sparse graph on n items to a sparse item–feature relation with m latent-labeled features. Cavity analysis predicted its detection boundary and Bayes-optimal BP behavior, leaving open a matching information-theoretic converse and polynomial-time achievability theorem. Let d_A, d_F denote the mean graph and item-side incidence degrees, s_A, s_F the corresponding one-edge label correlations, and $m/n \to \gamma$. For fixed known parameters away from criticality, we prove the sharp threshold $\Lambda := d_A s_A^2 + \frac{d_F^2 s_F^4}{\gamma} = 1$. For $\Lambda < 1$, the planted and null laws are mutually contiguous and weak recovery is impossible; for $\Lambda > 1$, they are asymptotically singular and randomized-polynomial-time weak recovery is possible. Thus two individually undetectable relations can jointly support efficient recovery, with no statistical–computational gap. The proof combines typed-cycle and self-avoiding-path expansions with a forest–parity reduction that controls overlaps sharing latent features, half-edges, or reversed segments. Large-scale BP experiments recover the predicted phase geometry and joint rescue, while weighted spectral inference also illustrates joint recovery.

1 INTRODUCTION

Community detection seeks to infer latent groups from interactions between their members (Abbe, 2018). The stochastic block model (SBM) (Holland et al., 1983) formalizes this problem by making edge probabilities depend on group membership. In the sparse regime considered here, each vertex has a bounded expected number of neighbors. Recovering the groups depends on whether these limited observations collectively carry enough information, a question captured by sharp recovery thresholds (Decelle et al., 2011b; Mossel et al., 2015; Massoulié, 2014; Mossel et al., 2018).

Many learning problems provide several sources of information about the same objects (Deshpande et al., 2018; Lu & Sen, 2023). Documents, for example, can link to one another and mention shared entities; both kinds of observations inform their organization (Chang & Blei, 2010; Balasubramanyan & Cohen, 2011). When both relations are sparse, each may be too weak to reveal communities on its own. Their combination can nevertheless carry useful information. This raises a basic question: *when do individually insufficient sparse relations jointly permit community recovery, and can this information be used in polynomial time?*

We study this question in the joint stochastic block model (JSBM), which couples a graph on item vertices with a bipartite graph between items and feature vertices. Both vertex populations have latent communities, and the two observed graphs share the same item labels. Thus evidence about an item comes both from its direct links and from its connections through latent feature communities. The cavity and belief-propagation (BP) analysis of Zhou et al. (2020) predicted an additive recovery boundary for this model, leaving open a matching information-theoretic converse and a polynomial-time achievability theorem.

We close this gap by establishing a sharp weak-recovery theory for community detection with sparse relational side information, providing a counterpart to sharp recovery results for the Gaussian contextual SBM (Lu & Sen, 2023). For the binary symmetric JSBM with fixed known parameters, we prove impossibility at and below the predicted boundary and randomized polynomial-time recovery above it. To our knowledge, this is the first sharp weak-recovery theorem for this model. The proof

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.

controls overlaps between sparse relational paths: reused incidence scores have label-dependent second moments of order m, and paths can share an incidence without sharing a complete feature wedge. Forest parity and typed overlap counting control these dependencies at logarithmic depth while preserving the threshold constant. The theorem establishes that two sparse relations, each insufficient for weak recovery, can jointly support recovery with no statistical–computational gap.

Concretely, let $V = [n]$ be item vertices and $U = [m]$ auxiliary feature vertices, with $m/n \to \gamma \in (0, \infty)$. Assign independent uniform signs $\sigma_i, \tau_\mu \in \{\pm 1\}$ to items and features.

Conditional on these labels, all observed edges are independent and

$$\mathbb{P}(A_{ij} = 1 \mid \sigma) = \begin{cases} a/n, & \sigma_i = \sigma_j, \\ b/n, & \sigma_i \neq \sigma_j, \end{cases} \qquad i < j, \tag{1}$$

$$\mathbb{P}(F_{i\mu} = 1 \mid \sigma, \tau) = \begin{cases} \alpha/m, & \sigma_i = \tau_\mu, \\ \beta/m, & \sigma_i \neq \tau_\mu, \end{cases} \qquad i \in V,\ \mu \in U. \tag{2}$$

Here a, b are the within- and between-class item-edge rates; α, β are the corresponding item–feature rates. Only the item adjacency matrix A and item–feature incidence matrix F are observed. Recovery targets σ up to global sign (Figure 1).

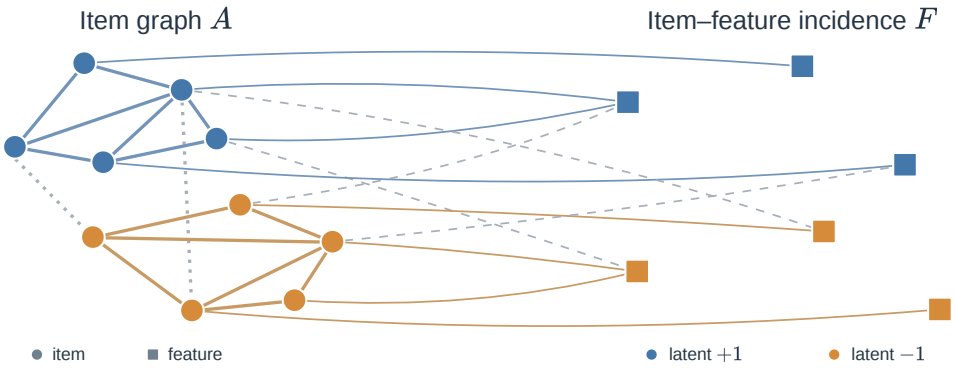

Figure 1: An assortative JSBM example. Circles denote items, squares features, and colors latent classes. Solid edges are within-class; dashed or dotted edges are between-class. Only A, F are observed; recovery targets item labels.

Set the item-side mean degrees $d_A = (a + b)/2$, $d_F = (\alpha + \beta)/2$ and the normalized channel eigenvalues $s_A = (a - b)/(a + b)$, $s_F = (\alpha - \beta)/(\alpha + \beta)$. Equivalently, the edge probabilities are $d_A(1 + s_A\sigma_i\sigma_j)/n$ and $d_F(1 + s_F\sigma_i\tau_\mu)/m$. An item edge transmits evidence in one step, whereas feature evidence traverses an item–feature–item wedge. The resulting effective strengths are

$$\Lambda_A = d_A s_A^2, \qquad \Lambda_F = \frac{d_F^2 s_F^4}{\gamma}. \tag{3}$$

The resulting signal-addition boundary is

$$\boxed{\Lambda := \Lambda_A + \Lambda_F = 1.} \tag{4}$$

When $\Lambda_A < 1$ and $\Lambda_F < 1$ but $\Lambda_A + \Lambda_F > 1$, neither view alone supports weak recovery, whereas their combination admits a polynomial-time estimator.

Key mechanism. Forest parity makes each path-pair moment depend only on repeated edges. Typed vertex counts balance their large sparse-incidence moments against vertex identifications. Summing these terms with channel-matched weights gives a sharp bound at logarithmic depth, including half-wedges and reversed segments.

Contributions.

1. **Sharp statistical–computational threshold.** We prove impossibility of weak recovery for $\Lambda \leq 1$ and randomized polynomial-time recovery for $\Lambda > 1$. The planted and null laws are mutually contiguous for $\Lambda < 1$ and asymptotically singular for $\Lambda > 1$, establishing the predicted boundary rigorously.
2. **Sharp overlap control via forest parity.** We derive an exact second-moment identity for overlapping relational paths (Equation (13)). Typed vertex counting then absorbs sparse-incidence factors, and channel-matched summation yields a sharp logarithmic-depth bound (Lemma 2). The bound covers partial wedges, shared features, and reversed segments.
3. **Joint recovery from individually insufficient relations.** The threshold identifies precisely when two subcritical channels jointly permit efficient recovery. It quantifies how evidence through latent features combines with direct graph evidence.

BP and weighted spectral experiments illustrate the predicted phase geometry and joint recovery at finite sizes. A paired-budget study across graph sizes examines recovery near the boundary and the effect of additional computation.

2 SHARP TRANSITION AND SIGNAL GEOMETRY

Writing $\theta = (d_A, d_F, s_A, s_F)$, let $\mathbb{P} = \mathbb{P}_{n,\theta}$ denote the planted joint law of (σ, τ, A, F) and, when comparing observation laws, its marginal on $\mathcal{O} = (A, F)$. Let $\mathbb{Q} = \mathbb{Q}_{n,\theta}$ be the degree-matched null law with mutually independent $A_{ij} \sim \mathrm{Bernoulli}(d_A/n)$ and $F_{i\mu} \sim \mathrm{Bernoulli}(d_F/m)$. Constants subscripted by θ may also depend on the fixed limiting aspect ratio γ. For an estimator $x = x(A, F) \in [-1, 1]^n$, define $\mathrm{ov}(x, \sigma) = n^{-1}|\langle x, \sigma\rangle|$; weak recovery means $\liminf_n \mathbb{E}_{\mathbb{P}} \mathrm{ov}(x, \sigma) > 0$. A test is consistent when the sum of its type-I and type-II errors tends to zero.

Theorem 1 (Sharp weak-recovery threshold and noncritical detection)**.** *Fix $d_A, d_F > 0$, $s_A, s_F \in (-1, 1)$, and let $m/n \to \gamma$. Assume these parameters are supplied to the estimator. With the channel strengths in* (3) *and Λ in* (4)*:*

1. *If $\Lambda \leq 1$, every estimator has expected overlap tending to zero. If $\Lambda < 1$, then $\mathbb{P}$ and $\mathbb{Q}$ are mutually contiguous.*
2. *If $\Lambda > 1$, then $\mathbb{P}$ and $\mathbb{Q}$ are asymptotically singular, and item-label weak recovery is achievable in randomized polynomial time.*

Weak recovery is therefore possible exactly when $\Lambda > 1$; the testing statements concern $\Lambda \neq 1$.

Joint recovery. For example, $\Lambda_A = \Lambda_F = .6$ puts each view below its individual weak-recovery threshold but their combination above the joint threshold. Theorem 1 gives a polynomial-time estimator from the joint observations. Section 6 studies finite-budget BP and weighted spectral inference in this regime.

When $s_F = 0$, the boundary is the ordinary binary-SBM Kesten–Stigum condition $d_A s_A^2 = 1$. When $s_A = 0$, it becomes the two-step bipartite condition $d_F^2 s_F^4/\gamma = 1$. Both assortative and disassortative channels enter through the squared correlations s_A^2, s_F^2.

Feature-mediated information. An item–feature edge has label correlation s_F, so the wedge $i \to \mu \to j$ has correlation s_F^2. An item has about d_F feature neighbors, each with about d_F/γ other item neighbors. The squared-channel branching contribution is

$$d_F \frac{d_F}{\gamma}(s_F^2)^2 = \frac{d_F^2 s_F^4}{\gamma} = \Lambda_F. \tag{5}$$

A direct item edge completes an item-to-item step and contributes $d_A s_A^2$. This path-level calculation explains both the different exponents and the aspect-ratio penalty. The proof therefore works on the incidence graph, retaining the shared-feature dependencies in $FF^\top$.

Observed vertex types. Item/feature types are observed, while each type contains two latent sign classes. The four-block representation separates known vertex types from the hidden item labels targeted by weak recovery. The null preserves the observed types and their edge densities.

Algorithm inputs and complexity. The estimator receives $\theta = (d_A, d_F, s_A, s_F)$ and reads n, m from the observations. For fixed $\Lambda > 1$, it uses $k = \lceil 4 \log n / \log \Lambda_n^{\mathrm{path}} \rceil$ for all large n, where $\Lambda_n^{\mathrm{path}} \to \Lambda$ includes the finite-size corrections in Section 4. Color coding costs $2^{O(k)}\,\mathrm{poly}(n) = n^{O(1+1/\log \Lambda)}$ arithmetic operations; rounding yields $\liminf_n \mathbb{E}\,\mathrm{ov} \geq c_\theta > 0$. The exponent increases as $\Lambda \downarrow 1$.

3 Related work

Networks with auxiliary information. Attributed-network methods combine links and vertex information (Bothorel et al., 2015). Representative approaches include joint latent-variable models (Zanghi et al., 2010; Xu et al., 2012), link–content learning (Yang et al., 2009; 2013), metadata-informed inference (Newman & Clauset, 2016), and graph augmentation (Zhou et al., 2009). Covariate-assisted spectral methods, joint criteria, and SDP (Binkiewicz et al., 2017; Zhang et al., 2016; Yan & Sarkar, 2021) establish benefits of fusion. Bipartite block models (Florescu & Perkins, 2016) and multilayer networks (Paul & Chen, 2020; Chatterjee et al., 2024) study multiple populations and networks.

Sharp thresholds and algorithms. Cavity predictions (Decelle et al., 2011b;a), SBM threshold proofs (Mossel et al., 2015; Massoulié, 2014; Mossel et al., 2018; Abbe & Sandon, 2016) and sparse spectral methods (Krzakala et al., 2013; Saade et al., 2014; Bordenave et al., 2018) provide theoretical and algorithmic foundations. Related analyses address mutual information at diverging degree (Deshpande et al., 2017) and dense-model contiguity (Banerjee, 2018). Path estimators and matrix rounding (Massoulié, 2014; Hopkins & Steurer, 2017; Lu & Sen, 2023) yield recovery algorithms.

Gaussian covariates and sparse relational features. The Gaussian contextual SBM (CSBM) observes a sparse graph and dense noisy covariates sharing a latent direction. Its transition was predicted by Deshpande et al. (2018) and proved at fixed degree by Lu & Sen (2023). JSBM observes sparse incidences to latent-labeled features, with a cavity/BP boundary predicted by Zhou et al. (2020). The path-based framework requires overlap bounds for repeated incidence scores with label-dependent second moments of order m (Equation (6)). Forest parity and typed counting provide the required overlap control, preserving the threshold constant.

Other informative relations. Weighted and edge-labeled models (Aicher et al., 2015; Xu et al., 2014; Lelarge et al., 2015) study informative pairwise observations. General SBM algorithms and weighted non-backtracking analyses give recovery guarantees for heterogeneous sparse graphs (Abbe & Sandon, 2016; Stephan & Massoulié, 2022). Planted-factor-graph theory relates recovery to testing away from criticality (Mossel et al., 2025). For JSBM, we establish matching impossibility and polynomial-time achievability for hidden item labels given the observed vertex types, including critical non-recovery.

4 From local signal addition to global recovery

We prove the recovery part of Theorem 1 using self-avoiding paths on the incidence graph. The main step is to bound the second moment of their weighted sum when paths share observations. Section 5 gives the non-recovery and testing arguments.

4.1 Sparse incidence moments and shared features

For an active feature channel, put $p_F = d_F/m$ and define the centered score $\widehat{F}_{i\mu} = (F_{i\mu} - p_F)/(p_F s_F)$. Its conditional mean is $\sigma_i \tau_\mu$, while its conditional second moment is exactly

$$\mathbb{E}[\widehat{F}_{i\mu}^2 \mid \sigma, \tau] = \frac{1 - p_F}{p_F s_F^2} + \frac{1 - 2p_F}{p_F s_F} \sigma_i \tau_\mu. \tag{6}$$

Both coefficients are of order m at fixed nonzero s_F. Thus the second moment of a path pair depends on which incidences it reuses and on the signs at their endpoints. We average the latent signs exactly and balance the surviving moments against the loss of free vertices, keeping item and feature identifications separate.

A feature induces several item–item wedges. Even conditional on all latent labels, those wedges need not be independent. In Figure 2, the two simple paths $\pi = (i, \mu, a, j)$ and $\pi' = (i, \mu, b, j)$ share only the incidence $i\mu$, not a complete feature wedge. If $q_x = \mathbb{P}(F_{x\mu} = 1 \mid \sigma, \tau)$, conditional edge independence gives

$$\mathrm{Cov}(F_{i\mu}F_{a\mu}, F_{i\mu}F_{b\mu} \mid \sigma, \tau) = q_i(1 - q_i)q_a q_b > 0. \tag{7}$$

We work on the original item–feature graph to track shared incidences. Centering the edge observations yields unbiased path scores; path-pair moments depend on shared edges.

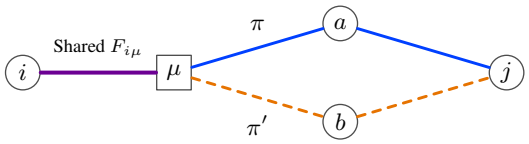

Figure 2: A half-wedge overlap: $\pi = i-\mu-a-j$ and $\pi' = i-\mu-b-j$ share $F_{i\mu}$ (purple), but no complete feature wedge. Circles are items; the square is a feature; a, b are distinct intermediate items. Blue and orange distinguish the two paths.

Overlap components. We group shared incidences into connected components. Vertex identifications balance their sparse moments, partial wedges contribute endpoint factors, and complete shared steps are summed with channel-matched weights. This componentwise accounting controls the second moment at logarithmic path lengths throughout $\Lambda > 1$.

4.2 The transfer operator

The two-type Kesten–Stigum matrix keeps the item and feature vertex types. With rows and columns ordered as (item, feature), define

$$K = \begin{pmatrix} d_A s_A^2 & d_F s_F^2 \\ d_F s_F^2/\gamma & 0 \end{pmatrix} = \begin{pmatrix} \Lambda_A & \sqrt{\gamma\Lambda_F} \\ \sqrt{\Lambda_F/\gamma} & 0 \end{pmatrix}. \tag{8}$$

Its spectral radius is

$$\rho(K) = \frac{\Lambda_A + \sqrt{\Lambda_A^2 + 4\Lambda_F}}{2}, \qquad \rho(K) > 1 \iff \Lambda_A + \Lambda_F > 1. \tag{9}$$

On the local two-type Galton–Watson limit, this squared-channel matrix determines the instability threshold for BP perturbations. Equation (9) turns a two-type branching instability into the additive boundary (4).

4.3 Controlling overlaps at logarithmic depth

We estimate the item-pair matrix $\sigma\sigma^\top$. The construction has three steps: match the path mixture to the channel strengths, control the second moment by overlap components, and convert matrix correlation to item-label recovery.

Typed paths and signal matching. For distinct items i, j, consider self-avoiding item-to-item paths of $k = \Theta_\theta(\log n)$ macro-steps, with a computable choice given in Appendix D. A macro-step is either one centered A edge or a two-edge centered F wedge. Use the incidence scores above and $\widehat{A}_{ij} = (A_{ij} - d_A/n)/(d_A s_A/n)$ on the active graph channel. For any self-avoiding path, conditional independence of its distinct edges makes the mean of the score product equal to the product of their endpoint signs. Every internal sign occurs twice and cancels, leaving exactly $\sigma_i\sigma_j$. This holds for

every step-type sequence, so a mixture whose weights sum to one retains the same signal. When both contributions are positive, the limiting variance-minimizing mixture is

$$\eta_A = \frac{\Lambda_A}{\Lambda}, \qquad \eta_F = \frac{\Lambda_F}{\Lambda}. \tag{10}$$

For any mixture $w_A + w_F = 1$, Cauchy–Schwarz gives $1 = (w_A + w_F)^2 \le (w_A^2/\Lambda_A + w_F^2/\Lambda_F)(\Lambda_A + \Lambda_F)$. Equality holds at $(w_A, w_F) = (\eta_A, \eta_F)$ in (10). Thus summing over the type of a complete shared macro-step attains the smallest possible limiting second-moment factor

$$\frac{\eta_A^2}{\Lambda_A} + \frac{\eta_F^2}{\Lambda_F} = \frac{1}{\Lambda}. \tag{11}$$

If one channel vanishes, step-type sequences using it and the corresponding term are omitted. The finite-size strengths are

$$\Lambda_{A,n}^{\mathrm{path}} = \frac{\Lambda_A}{1 - d_A/n}, \qquad \Lambda_{F,n}^{\mathrm{path}} = \frac{d_F^2 s_F^4 (n/m)}{(1 - d_F/m)^2}.$$

Set $\Lambda_n^{\mathrm{path}} = \Lambda_{A,n}^{\mathrm{path}} + \Lambda_{F,n}^{\mathrm{path}}$ and $\eta_{\chi,n} = \Lambda_{\chi,n}^{\mathrm{path}}/\Lambda_n^{\mathrm{path}}$ for $\chi \in \{A, F\}$. These converge to their limiting counterparts. For each step-type sequence, multiply the normalized edge scores along each path and average over all its self-avoiding embeddings from i to j. Mix these averages with the product of the corresponding finite-size step weights to obtain P_{ij}, and set $P_{ii} = 0$. Equation (11) makes Λ the quantity controlling the second moment of this matrix:

$$\mathbb{E}\|P\|_F^2 \le C_0 n^2, \qquad \mathbb{E}\langle P, \sigma\sigma^\top\rangle = n(n-1). \tag{12}$$

Forest parity and exact overlap moments. Let p_π be the product of normalized edge scores along a self-avoiding path π. For two such paths with the same endpoints, let e_A, e_F count their shared item edges and incidences. The edges used once have even degree at every vertex, so their latent signs cancel. The shared-edge graph is a forest: every nonempty subset has a leaf, whose independent sign averages to zero. Thus all label-dependent repeated-edge terms vanish, giving the exact identity

$$\boxed{\mathbb{E}[p_\pi p_{\pi'}] = \left(\frac{n - d_A}{d_A s_A^2}\right)^{e_A} \left(\frac{m - d_F}{d_F s_F^2}\right)^{e_F}.} \tag{13}$$

Only active channels enter this formula. It holds for every overlap pattern, including half-wedges and reversed segments (Lemma 5), reducing latent-label dependencies to explicit factors determined by the shared edges.

Two local examples show how vertex counting absorbs these factors. Put $z_F = d_F s_F^2$ and $c_{F,n} = 1 - d_F/m$ on an active feature channel. A repeated normalized incidence has constant second moment $c_{F,n} m / z_F$. In Figure 2, identifying the common feature μ costs a factor m^{-1} in the number of path pairs. Thus the shared half-wedge contributes

$$\underbrace{m^{-1}}_{\text{one shared feature}} \quad \underbrace{\frac{c_{F,n} m}{z_F}}_{\text{one repeated incidence}} = \frac{c_{F,n}}{z_F},$$

a bounded endpoint cost. An isolated interior shared wedge $b - \nu - c$ identifies two items and one feature. Its contribution before step-type weights is

$$n^{-2} m^{-1} \left(\frac{c_{F,n} m}{z_F}\right)^2 = \frac{1}{n \Lambda_{F,n}^{\mathrm{path}}}, \qquad \Lambda_{F,n}^{\mathrm{path}} = \frac{z_F^2 n}{c_{F,n}^2 m}. \tag{14}$$

These embedding ratios have a uniform $1 + O(k^2/n)$ falling-factorial correction. Longer interior components still pay one anchor loss n^{-1}; each complete shared step contributes its channel's inverse strength. Summing its two possible types with the squared step-type weights gives exactly $1/\Lambda_n^{\mathrm{path}}$, the finite-size version of (11). Partial wedges contribute bounded factors at component endpoints, while the complete steps sum geometrically with ratio $1/\Lambda_n^{\mathrm{path}}$.

Lemma 2 (Sharp incidence-overlap control). *Fix a parameter tuple with $\Lambda > 1$ and $k = O_\theta(\log n)$. For each fixed item pair $i \neq j$, exclude fully coincident paths. For the normalized, signal-weighted entry P_{ij}, the total second-moment contribution of all overlap diagrams with c_0 interior common-edge components and v_0 isolated shared nonendpoint vertices is at most*

$$C_\theta \left(\frac{C_\theta k^4}{n}\right)^{c_0} \left(\frac{C_\theta k^2}{n}\right)^{v_0}. \tag{15}$$

Here C_θ is finite and independent of n, k.

Counting mechanism. A common component is specified by its two endpoint positions in each path, an orientation, and finitely many half-wedge states: $O(k^4)$ choices. An isolated shared vertex needs only its two positions: $O(k^2)$ choices. Their n^{-1} penalties dominate these multiplicities at logarithmic depth. Each constrained macro-position receives its word weight once per path; complete shared steps therefore carry squared weights, whereas free positions sum to one. Partial-wedge weights enter only their endpoint factors. Boundary components are anchored at i or j; their lengths are summed geometrically using $\Lambda_n^{\mathrm{path}} > 1$. Fully coincident pairs contribute at most $Cn/(\Lambda_n^{\mathrm{path}})^k$, controlled by choosing the depth large enough. Together these bounds yield (12). Appendix D proves the lemma by reconstructing overlap diagrams from these records and illustrates the three overlap types in one example.

Polynomial-time evaluation and rounding. Color coding (Alon et al., 1995) assigns colors to the vertices of the expanded typed path (including auxiliary vertices) and evaluates the resulting subset dynamic program in $2^{O(k)}\,\mathrm{poly}(n) = n^{O(1)}$ time. A randomized dyadic semidefinite programming (SDP) projection followed by Gaussian sign rounding converts the matrix correlation in (12) into nonvanishing expected overlap; the conversion is proved in Appendix D.

5 Non-recovery and hypothesis testing

The remaining claims of Theorem 1 follow from likelihood-ratio moments, typed cycle counts, and a critical tree argument. Cycle counts characterize subcritical likelihood fluctuations and give a supercritical test. At the boundary, a two-type information contraction and a nonedge likelihood estimate establish non-recovery.

Below the boundary. Let $L_n = d\mathbb{P}/d\mathbb{Q}$. An exact Bernoulli replica identity followed by a uniformly integrable quadratic expansion gives

$$\mathbb{E}_{\mathbb{Q}} L_n^2 \longrightarrow \frac{\exp\{-\Lambda_A/2 - \Lambda_A^2/4 - \Lambda_F/2\}}{\sqrt{1-\Lambda_A-\Lambda_F}}. \tag{16}$$

This is finite precisely for $\Lambda < 1$. Fixing finitely many item labels in the same calculation makes the corresponding conditional observation laws merge, forcing posterior pair correlations and every estimator's weak overlap to vanish. Typed-cycle small-subgraph conditioning supplies the reverse contiguity. Appendix B proves the second-moment and posterior steps, and Appendix C proves the cycle likelihood-ratio limit. At $\Lambda = 1$, critical two-type tree contraction and a boundary-reveal tree-to-graph comparison rule out weak recovery (Appendix B.1).

Above the boundary. Set $\gamma_n = m/n$ and $\Lambda_n^{\mathrm{cyc}} = d_A s_A^2 + d_F^2 s_F^4/\gamma_n$. For $\ell \to \infty$ slowly, the aggregate $T_{\ell,n}$ of centered typed cycles, weighted using these finite-size parameters, satisfies

$$\mathbb{E}_{\mathbb{P}} T_{\ell,n} = (1+o(1))\sqrt{\frac{(\Lambda_n^{\mathrm{cyc}})^\ell}{2\ell}}, \qquad \mathrm{Var}_{\mathbb{P}}(T_{\ell,n}) = 1 + o(1). \tag{17}$$

Under the null, $\mathbb{E}_{\mathbb{Q}} T_{\ell,n} = 0$ and $\mathrm{Var}_{\mathbb{Q}}(T_{\ell,n}) = 1$. The quotient union of every proper cycle overlap has positive edge–vertex excess, so the covariance correction vanishes. Since $\Lambda_n^{\mathrm{cyc}} \to \Lambda > 1$, the planted mean diverges, and Chebyshev's inequality gives a consistent test and hence asymptotic singularity; Appendix C gives the enumeration. Randomized polynomial-time weak recovery is completed by the typed-SAW, color-coding, and rounding construction of Section 4 and Appendix D.

6 NUMERICAL ILLUSTRATIONS

The recovery theorem uses the self-avoiding-path construction. We evaluate BP and weighted Bethe–Hessian (BH) on additive phase geometry and recovery from individually subcritical views. Both receive the observed graphs and generating parameters. BP summaries include the terminal output of every run, including those reaching the iteration cap.

Phase geometry. Figure 3 plots the soft order parameter $|n^{-1}\sum_i \sigma_i \tanh(h_i^{\mathrm{BP}})|$ in two 101×101 phase planes and along three cuts. The two planes are independent scans at different fixed degrees. The additive boundary is a straight line in contribution coordinates (Λ_A, Λ_F) but a curve in the edge-rate ratios $\epsilon_A = b/a$ and $\epsilon_F = \beta/\alpha$. The cuts fix $\Lambda_A/\Lambda \in \{.25, .50, .75\}$. Throughout the cuts, both individual strengths remain below one, while BP overlap rises near the joint boundary, illustrating signal addition. In these cuts, 15.9% of runs reach the iteration cap, concentrated near the boundary. Appendix E reports convergence diagnostics and parameter controls.

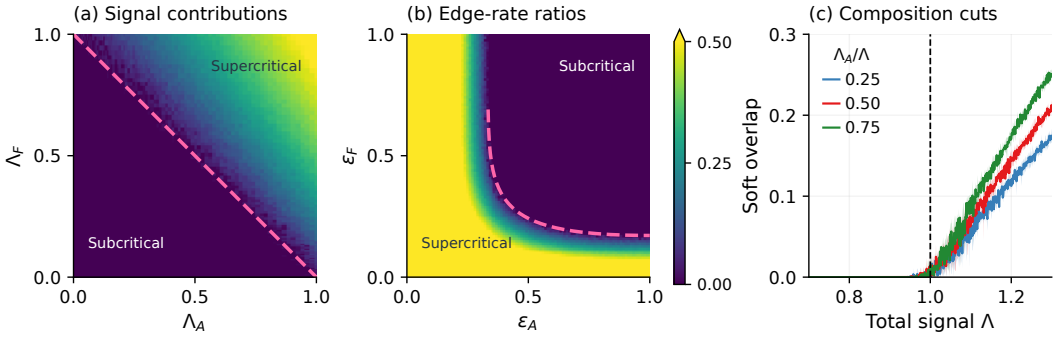

Figure 3: BP phase geometry at $n = m = 200{,}000$; all panels show soft overlap. **(a)** Signal contributions, $d_A = d_F = 3$. **(b)** Edge-rate ratios, $d_A = 4, d_F = 2$. Heatmaps use one instance per point and a shared scale capped at .5; dashed curves mark $\Lambda = 1$. The $\epsilon = 0$ edges additionally show noiseless channels outside the theorem's $|s| < 1$ assumptions. **(c)** Fixed-composition cuts, $d_A = d_F = 3$: mean ± 2 SE over five instances per point.

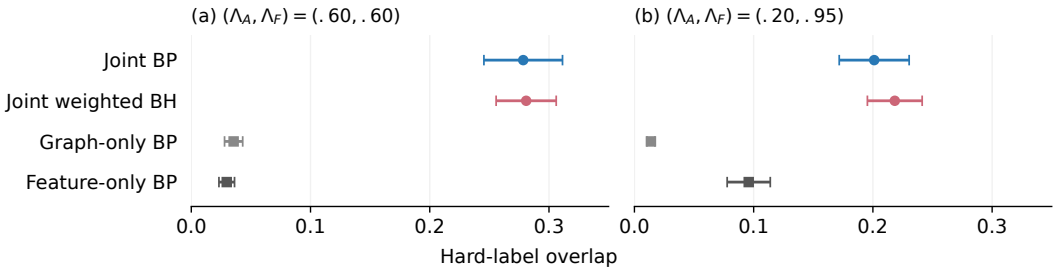

Figure 4: Joint recovery from individually subcritical views. At $n = m = 5000$, $d_A = d_F = 3$, points and bars show hard-label overlap $n^{-1}|\langle \widehat{\sigma}, \sigma \rangle|$, mean ± 2 SE over 50 paired instances.

Joint recovery. Figure 4 compares joint BP and weighted BH with single-view BP on the same graphs. At $(\Lambda_A, \Lambda_F) = (.60, .60)$, joint BP reaches $.278 \pm .033$ hard-label overlap, versus $.036 \pm .008$ for graph-only and $.030 \pm .007$ for feature-only BP. At $(.20, .95)$, joint BP reaches $.201 \pm .029$, versus $.096 \pm .018$ for feature-only BP. Weighted BH reaches $.281 \pm .025$ and $.218 \pm .023$, respectively. Both methods improve on single-view BP in these settings. Appendix E gives inference equations, paired protocols, and additional fusion/projection comparisons.

Finite-size behavior near the boundary. Figure 5 uses 1,400 independent graphs: four sizes from 5,000 to 200,000, seven strengths near $\Lambda = 1$, and 50 graphs per coordinate, with $\Lambda_A = \Lambda_F = \Lambda/2$. Under the 5,600-iteration budget, mean soft overlap at $\Lambda = .98$ falls from .0172 to .00120 as size

increases, while at $\Lambda = 1.10$ it stays between .06457 and .07395. Pairing each run with a 1,400-iteration budget on the same graph and initialization isolates the effect of additional computation. At $n = 200{,}000$, the mean change at $\Lambda = 1.10$ is below 3×10^{-7} in magnitude and all longer runs converge; at $\Lambda = 1.02$, 20/50 still reach the longer cap. Appendix E.3.2 reports both soft and hard overlaps, error bars, and stopping frequencies across all sizes.

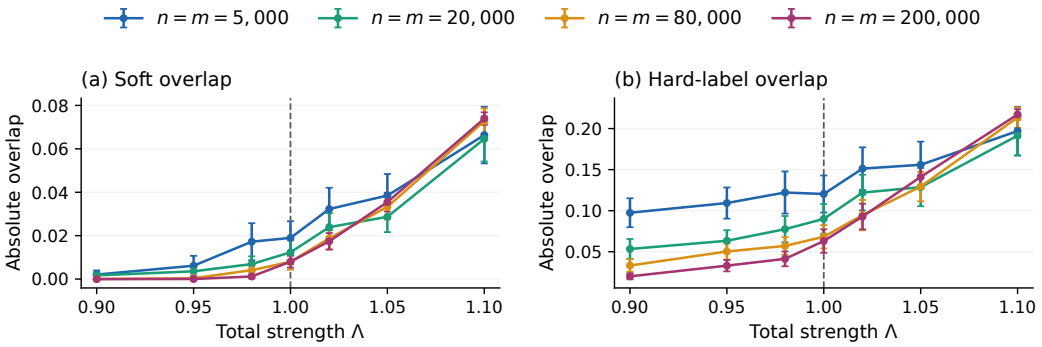

Figure 5: Finite-size behavior near the additive boundary, with $\Lambda_A = \Lambda_F = \Lambda/2$ and $d_A = d_F = 3$. Points and bars show mean ± 2 SE over 50 independent graphs per coordinate, using a 5,600-iteration cap. Dashed lines mark $\Lambda = 1$.

7 Conclusion

Two sparse relations can each be insufficient for community recovery yet jointly permit it. For the balanced binary JSBM with fixed known parameters, we establish that polynomial-time item weak recovery is possible exactly when $d_A s_A^2 + d_F^2 s_F^4/\gamma > 1$, and locate the noncritical testing boundary. The paired BP and weighted-spectral experiments illustrate the joint-recovery regime at finite size. Learning channel correlations, resolving critical testing, and extending the result beyond two communities remain open directions.

Reproducibility Statement

Section 2 states the model and theorem assumptions, and Appendix A collects the notation. Appendix B details the second-moment bounds and weak-recovery impossibility, including the critical point. Appendix C derives the typed-cycle likelihood-ratio limit and the supercritical test. Appendix D develops the recovery construction, including forest parity, overlap enumeration, path-pair moment bounds, color coding, and matrix-to-label rounding. Appendix E specifies the inference equations, initialization, stopping rules, and evaluation protocols for the numerical illustrations.

AI Use Statement

Generative AI tools assisted literature search and reference verification, drafting and revising manuscript passages, and language polishing. They provided feedback on research methodology and experimental protocols, assisted implementation of inference methods and code for generating synthetic JSBM graphs and labels, and helped prepare plots and interpret numerical results. AI tools also supported correctness checks of existing mathematical arguments and revisions to their exposition. Further uses included pre-submission checks and simulated reviewer feedback.

The authors reviewed the AI-assisted text, checked plotted values against recorded experimental results, and tested the relevant code components. The authors take responsibility for the final text, mathematical claims and proofs, experimental results, figures, and code.

REFERENCES

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018. URL https://jmlr.org/papers/v18/16-480.html.

Emmanuel Abbe and Colin Sandon. Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, 2016.

C. Aicher, A. Z. Jacobs, and A. Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248, 2015.

Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *Journal of the ACM*, 42(4):844–856, 1995.

Ramnath Balasubramanyan and William W. Cohen. Block-LDA: Jointly modeling entity-annotated text and entity-entity links. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pp. 450–461, 2011.

Debapratim Banerjee. Contiguity and non-reconstruction results for planted partition models: the dense case. *Electronic Journal of Probability*, 23(18):1–28, 2018.

Norbert Binkiewicz, Joshua T. Vogelstein, and Karl Rohe. Covariate-assisted spectral clustering. *Biometrika*, 104(2):361–377, 2017.

Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs. *The Annals of Probability*, 46 (1):1–71, 2018.

Cécile Bothorel, Juan David Cruz, Matteo Magnani, and Barbora Micenková. Clustering attributed graphs: Models, measures and methods. *Network Science*, 3(3):408–444, 2015.

Jonathan Chang and David M. Blei. Hierarchical relational models for document networks. *The Annals of Applied Statistics*, 4(1):124–150, 2010.

Anirban Chatterjee, Sagnik Nandy, and Ritwik Sadhu. Detecting planted partition in sparse multilayer networks. *Information and Inference: A Journal of the IMA*, 13(3):iaae019, 2024.

Aurélien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011a.

Aurélien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6): 065701, 2011b.

Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2): 125–170, 2017.

Yash Deshpande, Andrea Montanari, Elchanan Mossel, and Subhabrata Sen. Contextual stochastic block models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, 2018.

Laura Florescu and Will Perkins. Spectral thresholds in the bipartite stochastic block model. In *Proceedings of the Conference on Learning Theory (COLT)*, volume 49 of *Proceedings of Machine Learning Research*, pp. 943–959, 2016.

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

Samuel B. Hopkins and David Steurer. Efficient Bayesian estimation from few samples: Community detection and related problems. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 379–390, 2017.

Svante Janson. Random regular graphs: Asymptotic distributions and contiguity. *Combinatorics, Probability and Computing*, 4(4):369–405, 1995.

Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. Wiley, 2000.

Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

Marc Lelarge, Laurent Massoulié, and Jiaming Xu. Reconstruction in the labelled stochastic block model. *IEEE Transactions on Network Science and Engineering*, 2(4):152–163, 2015.

Chen Lu and Subhabrata Sen. Contextual stochastic block model: Sharp thresholds and contiguity. *Journal of Machine Learning Research*, 24(54):1–34, 2023.

Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, pp. 694–703, 2014.

Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162:431–461, 2015.

Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

Elchanan Mossel, Allan Sly, and Youngtak Sohn. Weak recovery, hypothesis testing, and mutual information in stochastic block models and planted factor graphs. In *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, pp. 2062–2073, 2025.

M. E. J. Newman and Aaron Clauset. Structure and inference in annotated networks. *Nature Communications*, 7(1):11863, 2016.

Subhadeep Paul and Yuguo Chen. Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *The Annals of Statistics*, 48(1):230–250, 2020.

Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral clustering of graphs with the Bethe Hessian. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 27, 2014.

Ludovic Stephan and Laurent Massoulié. Non-backtracking spectra of weighted inhomogeneous random graphs. *Mathematical Statistics and Learning*, 5(3/4):201–271, 2022.

Jiaming Xu, Laurent Massoulié, and Marc Lelarge. Edge label inference in generalized stochastic block models: From spectral theory to impossibility results. In *Proceedings of the Conference on Learning Theory (COLT)*, volume 35 of *Proceedings of Machine Learning Research*, pp. 903–920, 2014.

Zhiqiang Xu, Yiping Ke, Yi Wang, Hong Cheng, and James Cheng. A model-based approach to attributed graph clustering. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pp. 505–516, 2012.

Bowei Yan and Purnamrita Sarkar. Covariate regularized community detection in sparse graphs. *Journal of the American Statistical Association*, 116(534):734–745, 2021.

Jaewon Yang, Julian McAuley, and Jure Leskovec. Community detection in networks with node attributes. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, pp. 1151–1156, 2013.

Tianbao Yang, Rong Jin, Yun Chi, and Shenghuo Zhu. Combining link and content for community detection: A discriminative approach. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 927–936, 2009.

Hugo Zanghi, Stevenn Volant, and Christophe Ambroise. Clustering based on random graph model embedding vertex features. *Pattern Recognition Letters*, 31(9):830–836, 2010.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. Community detection in networks with node features. *Electronic Journal of Statistics*, 10(2):3153–3178, 2016.

Pengfei Zhou, Tianyi Li, and Pan Zhang. Phase transitions and optimal algorithms for semisupervised classifications on graphs: From belief propagation to graph convolution network. *Physical Review Research*, 2(3):033325, 2020.

Yang Zhou, Hong Cheng, and Jeffrey Xu Yu. Graph clustering based on structural/attribute similarities. *Proceedings of the VLDB Endowment*, 2(1):718–729, 2009.

A NOTATION

Throughout the appendix, we use

$$\Lambda_A = d_A s_A^2, \qquad \Lambda_F = \frac{d_F^2 s_F^4}{\gamma}, \qquad \Lambda = \Lambda_A + \Lambda_F.$$

The within/between-class rates (a, b, α, β) are defined in Section 1. The asymmetric powers in Λ_A and Λ_F reflect the one-edge graph step and the two-edge feature wedge. All asymptotic statements hold along sequences with $m/n \to \gamma$ and fixed $d_A, d_F > 0$ and $s_A, s_F \in (-1, 1)$. Detection always refers to testing the marginal observation laws $\mathbb{P}$ and $\mathbb{Q}$; weak recovery targets the item labels and is invariant to their global sign.

Roadmap. Section B proves the subcritical likelihood-ratio bound and impossibility of weak recovery, including the critical case in Section B.1; Section C supplies reverse contiguity and the supercritical detection test; and Section D constructs polynomial-time weak recovery throughout the supercritical region. Section E then records the experimental protocol and additional controls.

The transfer matrix, its Perron root, and the one-channel limits are given in Section 4.2 and Section 2. The finite-size likelihood-ratio and path strengths agree: $\Lambda_{\chi,n}^{\mathrm{LR}} = \Lambda_{\chi,n}^{\mathrm{path}}$ for $\chi \in \{A, F\}$. Cycle strengths omit the Bernoulli factors: $\Lambda_{A,n}^{\mathrm{cyc}} = \Lambda_A$ and $\Lambda_{F,n}^{\mathrm{cyc}} = d_F^2 s_F^4 (n/m)$. Their limits agree in all three uses.

Background. For random-graph counts, multitype branching, and cavity/BP methods, see Janson et al. (2000), Bollobás et al. (2007), and Mézard & Montanari (2009), respectively.

B SECOND MOMENTS AND RECOVERY IMPOSSIBILITY

Let $L_n = d\mathbb{P}/d\mathbb{Q}$. For $X \sim \mathrm{Bernoulli}(q)$ and

$$\ell_p(X) = \left(\frac{p}{q}\right)^X \left(\frac{1-p}{1-q}\right)^{1-X},$$

direct summation gives

$$\mathbb{E}_q[\ell_{p_1}(X)\ell_{p_2}(X)] = 1 + \frac{(p_1 - q)(p_2 - q)}{q(1-q)}. \tag{18}$$

Take two independent latent-label replicas and write $r_i = \sigma_i \sigma_i'$, $t_\mu = \tau_\mu \tau_\mu'$, together with

$$X_n = n^{-1/2} \sum_i r_i, \qquad Y_m = m^{-1/2} \sum_\mu t_\mu.$$

Applying (18) entrywise produces the exact formula

$$\mathbb{E}_{\mathbb{Q}} L_n^2 = \mathbb{E}_{r,t} \prod_{i<j} \left(1 + \frac{\Lambda_{A,n}^{\mathrm{LR}}}{n} r_i r_j\right) \prod_{i,\mu} \left(1 + \sqrt{\frac{\Lambda_{F,n}^{\mathrm{LR}}}{nm}}\, r_i t_\mu\right), \tag{19}$$

with coefficients

$$\Lambda_{A,n}^{\mathrm{LR}} = \frac{\Lambda_A}{1 - d_A/n}, \qquad \Lambda_{F,n}^{\mathrm{LR}} = \frac{d_F^2 s_F^4 (n/m)}{(1 - d_F/m)^2}.$$

Thus $\Lambda_{A,n}^{\mathrm{LR}} \to \Lambda_A$ and $\Lambda_{F,n}^{\mathrm{LR}} \to \Lambda_F$; the square root in (19) is the finite-n feature amplitude. Uniformly over the replicas, Taylor expansion of the logarithm in (19) gives

$$\begin{aligned} \log(\text{integrand}) = {} & \frac{\Lambda_{A,n}^{\mathrm{LR}}}{2} X_n^2 + \sqrt{\Lambda_{F,n}^{\mathrm{LR}}}\, X_n Y_m - \frac{\Lambda_{A,n}^{\mathrm{LR}}}{2} - \frac{(\Lambda_{A,n}^{\mathrm{LR}})^2}{4}\left(1 - \frac{1}{n}\right) \\ & - \frac{\Lambda_{F,n}^{\mathrm{LR}}}{2} + O(n^{-1}). \end{aligned} \tag{20}$$

The cubic remainder is $O(n^{-1})$ because every factor is $1 + O(n^{-1})$ and $m \asymp n$.

The central limit theorem sends (X_n, Y_m) to two independent standard Gaussians. The passage through the exponential is justified uniformly: if $\Lambda < 1$, choose $\delta > 0$ so that, eventually,

$$(1+\delta)\Lambda_{A,n}^{\mathrm{LR}} + (1+\delta)^2 \Lambda_{F,n}^{\mathrm{LR}} < 1.$$

Conditioning on X_n, Hoeffding's lemma for Y_m, followed by the Gaussian representation of a Rademacher quadratic exponential, bounds the $(1+\delta)$ moment by a fixed multiple of $(1 - q_n)^{-1/2}$, where the displayed left side is $q_n < 1$. The integrands are therefore uniformly integrable. Evaluating the limiting Gaussian integral proves

$$\mathbb{E}_{\mathbb{Q}} L_n^2 \longrightarrow \frac{\exp\{-\Lambda_A/2 - \Lambda_A^2/4 - \Lambda_F/2\}}{\sqrt{1 - \Lambda_A - \Lambda_F}}. \tag{21}$$

Cauchy–Schwarz now shows that $\mathbb{P}$ is contiguous to $\mathbb{Q}$. The reverse direction of contiguity is established by the typed-cycle likelihood-ratio limit in Section C; the remainder of this section establishes the separate weak-recovery impossibility statement.

To establish recovery impossibility, fix r item labels and let $L_{n,s}$ be the likelihood ratio conditional on their sign vector $s \in \{\pm 1\}^r$. In the replica calculation the corresponding r_i are deterministic; this displaces X_n by only $O(n^{-1/2})$. The same uniform-integrability argument gives, for every fixed s, t,

$$\mathbb{E}_{\mathbb{Q}} L_{n,s} L_{n,t} \longrightarrow \frac{\exp\{-\Lambda_A/2 - \Lambda_A^2/4 - \Lambda_F/2\}}{\sqrt{1 - \Lambda}}.$$

Consequently $\|L_{n,s} - L_{n,t}\|_2 \to 0$ and the corresponding conditional observation laws merge in total variation. In particular, for fixed $i \neq j$,

$$\mathbb{E}\,|\mathbb{E}[\sigma_i \sigma_j \mid A, F]| \to 0. \tag{22}$$

By Bayes' rule, the posterior probabilities of (σ_i, σ_j) are proportional to the four conditional likelihood ratios. Pairwise TV merging of the conditional observation laws implies convergence of this posterior in mean TV to the uniform prior, giving (22).

Now let $\widehat{\sigma} \in [-1, 1]^n$ be any estimator, possibly using an independent random seed Z, and put $R_n = n^{-1} \sum_i \sigma_i \widehat{\sigma}_i$. Conditional on $\mathcal{O} = (A, F)$, define $q_{ij}(\mathcal{O}) = \mathbb{E}_Z[\widehat{\sigma}_i \widehat{\sigma}_j \mid \mathcal{O}]$. Since $|q_{ij}| \leq 1$,

$$\begin{aligned} \mathbb{E} R_n^2 &\leq \frac{1}{n} + \frac{1}{n^2} \sum_{i \neq j} \mathbb{E}\,|\mathbb{E}[\sigma_i \sigma_j \mid \mathcal{O}]| \\ &= \frac{1}{n} + \left(1 - \frac{1}{n}\right) \mathbb{E}\,|\mathbb{E}[\sigma_1 \sigma_2 \mid \mathcal{O}]| = o(1), \end{aligned}$$

where the equality uses model exchangeability and allows arbitrary estimators. Thus $\mathbb{E}\,\mathrm{ov}(\widehat{\sigma}, \sigma) = \mathbb{E}|R_n| \to 0$, proving the subcritical recovery claim.

B.1 WEAK RECOVERY AT THE CRITICAL POINT

Proposition 3 (Critical non-recovery)**.** *Under the assumptions of Theorem 1, if* $\Lambda = 1$*, then* $\mathbb{E}|\mathbb{E}[\sigma_i\sigma_j \mid A, F]| \to 0$ *for distinct fixed items* i, j*. Consequently every estimator in* $[-1, 1]^n$ *has expected overlap tending to zero.*

Proof. Following the tree-to-graph argument of Mossel et al. (2015, Section 4), we establish the two-type critical contraction and bound the nonedge likelihood.

Critical tree contraction. Let T be the two-type Poisson Galton–Watson tree with independent offspring counts and mean offspring matrix

$$M_0 = \begin{pmatrix} d_A & d_F \\ d_F/\gamma & 0 \end{pmatrix}.$$

Generate its topology independently of a uniform root sign, then broadcast signs with correlation s_A on item edges and s_F on incidence edges. For root type $a \in \{I, F\}$, let $J_{a,t} = I(\xi_\rho; \xi_{\partial T_t} \mid T)$, with information in nats and ∂T_t the vertices at depth t. Conditional information here averages over T. Data processing on the same underlying tree makes $J_{a,t}$ nonincreasing in t. For a uniform sign X and an observation Y, put $u = \mathbb{E}[X \mid Y]$. Then $I(X;Y) = \mathbb{E}f(u)$, where

$$f(u) = \frac{1+u}{2}\log(1+u) + \frac{1-u}{2}\log(1-u) = \sum_{k\geq 1} \frac{u^{2k}}{2k(2k-1)}.$$

Thus $u^2/2 \leq f(u) \leq (\log 2)u^2$ and, for $|s| < 1$,

$$f(su) \leq s^2 f(u) - \frac{s^2(1-s^2)}{12}u^4.$$

When Y observes a child's subtree, a parent–child channel of correlation s gives parent posterior mean su. Jensen's inequality bounds its information by $s^2 I(X;Y) - c(s)I(X;Y)^2$, where $c(s) = s^2(1-s^2)/\{12(\log 2)^2\}$. Child observations are independent conditional on the parent sign and T, so their joint information is at most the sum of their individual information. Averaging over the independent offspring and subtrees gives, with $c_A = c(s_A)$ and $c_F = c(s_F)$,

$$\begin{aligned} J_{I,t+1} &\leq \Lambda_A J_{I,t} + d_F s_F^2 J_{F,t} - d_A c_A J_{I,t}^2 - d_F c_F J_{F,t}^2, \\ J_{F,t+1} &\leq (d_F/\gamma)s_F^2 J_{I,t} - (d_F/\gamma)c_F J_{I,t}^2. \end{aligned}$$

If $s_F \neq 0$, then K is irreducible and, at $\Lambda = 1$, $w = (1, d_F s_F^2)^\top > 0$ satisfies $w^\top K = w^\top$. Pass to the nonnegative limits of $J_{I,t}, J_{F,t}$ and multiply these inequalities by $w^\top$. The linear terms cancel; since $c_F > 0$, both limits must be zero. If $s_F = 0$, then $\Lambda_A = 1$ and $c_A > 0$; the first inequality alone forces $J_{I,t} \to 0$. Writing $h_t = \mathbb{E}[\xi_\rho \mid T, \xi_{\partial T_t}]$ for an item root, we have in either case

$$\mathbb{E}h_t^2 \leq 2J_{I,t} \longrightarrow 0. \tag{23}$$

Transfer to the observed graph. Write G for the observed typed incidence graph and $\xi = (\sigma, \tau)$. Fix an integer $r \geq 1$, let B_r be the radius-r ball about item 1, and set $S = B_{r-1}$, $D = B_r \setminus S$, $W = (V \cup U) \setminus B_r$. For fixed r, counting paths gives $\mathbb{E}|B_r| \leq C_{\theta,r}$, and the labeled ball converges to the depth-r broadcast tree above. Indeed, the two populations' empirical sign proportions tend to $1/2$; exploration gives independent Poisson offspring in the limit, with same/opposite-sign probabilities $(1 \pm s_A)/2$ or $(1 \pm s_F)/2$. Stopping exploration at L vertices bounds collision probability by $O_\theta(L^2/n)$. First letting $n \to \infty$, then $L \to \infty$, proves fixed-radius convergence for $m/n \to \gamma$.

The sets S, D, W are determined by G. Reveal all signs on S^c and, conditional on (G, ξ_{S^c}), compare the law on S with the edge-only tree law given ξ_D. For a vertex set H, write $M_{H,I} = \sum_{v\in H\cap V} \xi_v$ and $M_{H,F} = \sum_{v\in H\cap U} \xi_v$. There are no observed edges from S to W. On $|B_r| \leq L$, their nonedge log-likelihood, uniformly over candidate signs on S, equals a sign-independent constant plus

$$-\frac{d_A s_A}{n} M_{S,I} M_{W,I} - \frac{d_F s_F}{m}(M_{S,I} M_{W,F} + M_{S,F} M_{W,I}) + O_\theta(L/n). \tag{24}$$

The error follows by summing $O_\theta(n^{-2})$ Taylor remainders over $O(Ln)$ pairs. The true total sign sums have magnitudes at most $n^{2/3}$ and $m^{2/3}$ with probability tending to one. Removing B_r

changes either sum by at most L, so the displayed sign-dependent terms are $O_\theta(Ln^{-1/3} + L^2/n) = o(1)$. This estimate holds uniformly over candidate signs on S whenever the revealed outside sums satisfy these bounds. Nonedges inside B_r involving S contribute only $O_\theta(L^2/n)$. All remaining likelihood factors depend only on revealed signs and cancel. Let π_0 be the law proportional to $\prod_{uv\in E(B_r)}(1 + s_{uv}\xi_u\xi_v)$ with ξ_D fixed. The posterior has the form $\pi(z) = \pi_0(z)e^{\varepsilon(z)}/\mathbb{E}_{\pi_0}e^{\varepsilon}$, where $\sup_z |\varepsilon(z)| \le \delta_{n,L} = O_\theta(Ln^{-1/3} + L^2/n)$ after removing a constant. Thus $\mathrm{TV}(\pi, \pi_0) \le (e^{2\delta_{n,L}} - 1)/2 = o(1)$ for fixed L. On the tree event π_0 is exactly the broadcast posterior. The estimates are uniform on the stated events; boundedness and tightness let $L \to \infty$. Consequently, for $H_{n,r} = \mathbb{E}[\sigma_1 \mid G, \xi_{S^c}]$,

$$\lim_{n\to\infty} \mathbb{E}H_{n,r}^2 = \mathbb{E}h_r^2. \tag{25}$$

Finally let $q_{12}(G) = \mathbb{E}[\sigma_1\sigma_2 \mid G]$. Global sign symmetry gives $\mathbb{E}[\sigma_1 \mid G, \sigma_2] = \sigma_2 q_{12}(G)$. Put $H'_{n,r} = \mathbb{E}[\sigma_1 \mid G, \xi_{S^c}, \sigma_2]$. Conditional Jensen gives $\mathbb{E}q_{12}^2 \le \mathbb{E}(H'_{n,r})^2$. On the G-measurable event $2 \notin S$, $H'_{n,r} = H_{n,r}$; on its complement both are bounded by one. Hence $\mathbb{E}(H'_{n,r})^2 \le \mathbb{E}H_{n,r}^2 + \mathbb{P}(2 \in S)$, with $\mathbb{P}(2 \in S) = (\mathbb{E}|S \cap V| - 1)/(n-1) = O_{\theta,r}(n^{-1})$. For every fixed r,

$$\limsup_{n\to\infty} \mathbb{E}q_{12}(G)^2 \le \mathbb{E}h_r^2 \le 2J_{I,r}.$$

Let $r \to \infty$ and apply Cauchy–Schwarz to obtain (22). Exchangeability gives the same result for any distinct fixed items. The overlap bound preceding this proposition, which also allows independent estimator randomness, completes the proof. □

C TYPED CYCLES AND THE LIKELIHOOD-RATIO LIMIT

A word $w \in \{A, F\}^\ell$ encodes a simple cycle on ℓ item vertices: an A letter is a direct item edge and an F letter is a wedge through a new feature. Let $r(w)$ and $q(w)$ be its numbers of A and F letters. Words are identified under rotation and reversal, and $h(w)$ denotes the stabilizer size. At length two, AF and FF are admissible; AA repeats one undirected edge and is excluded. There are no length-one cycles. Define

$$\mu_w = \frac{d_A^{r(w)}(d_F^2/\gamma)^{q(w)}}{h(w)}, \qquad \delta_w = s_A^{r(w)}s_F^{2q(w)}. \tag{26}$$

For any fixed family of words, the corresponding unrooted simple-cycle counts converge under $\mathbb{Q}$ to independent $\mathrm{Pois}(\mu_w)$ variables. This follows from joint factorial moments. Vertex-disjoint embeddings give the product of the means. If two distinct cycle objects overlap, their quotient union has cycle rank at least two and hence $|E| - |V| \ge 1$ within that connected component. Since $m \asymp n$, its total contribution is $O(n^{-1})$. Under $\mathbb{P}$, averaging over the labels leaves only the empty and full even edge sets of each simple cycle, so its mean is multiplied by $1 + \delta_w$; the same excess argument removes overlapping families. Therefore the planted limits are independent $\mathrm{Pois}(\mu_w(1 + \delta_w))$ variables.

Orbit–stabilizer gives, for $\ell \ge 3$,

$$\sum_{w:\,|w|=\ell} \mu_w\delta_w^2 = \frac{\Lambda^\ell}{2\ell}. \tag{27}$$

At length two one subtracts the inadmissible AA mass $\Lambda_A^2/4$. Hence

$$S := \sum_w \mu_w\delta_w^2 = -\frac{1}{2}\log(1 - \Lambda) - \frac{\Lambda}{2} - \frac{\Lambda_A^2}{4} < \infty \tag{28}$$

exactly when $\Lambda < 1$. Define

$$L_\infty = \prod_w (1 + \delta_w)^{X_w} e^{-\mu_w\delta_w}. \tag{29}$$

Its finite restrictions

$$L_J = \prod_{w\in J} (1 + \delta_w)^{X_w} e^{-\mu_w\delta_w}$$

are mean-one L^2 martingales and $\mathbb{E}L_J^2 = \exp(\sum_{w\in J}\mu_w\delta_w^2)$. Thus they converge in L^2 to the product in (29). It is strictly positive almost surely: after finitely many terms, the centered logarithmic series has summable variance and the deterministic remainder is bounded by a constant times (28).

We approximate L_n by functions of finitely many bounded cycle counts. Let $C_{n,J}$ be the vector of counts for a finite word set J, and let $l_J(x) = \prod_{w\in J}(1+\delta_w)^{x_w}e^{-\mu_w\delta_w}$. The limiting count law under $\mathbb{P}$ has density l_J relative to its limiting law under $\mathbb{Q}$. For the finite box $D_M = \{0,\dots,M\}^J$ put $g_{J,M} = l_J\mathbf{1}_{D_M}$ and $a_{J,M} = \mathbb{E}[g_{J,M}(X_J)^2]$, where X_J has the null Poisson law. Convergence of the finitely many atom probabilities under both laws gives

$$\mathbb{E}_{\mathbb{Q}}g_{J,M}(C_{n,J})^2 \to a_{J,M}, \qquad \mathbb{E}_{\mathbb{Q}}[L_n g_{J,M}(C_{n,J})] = \mathbb{E}_{\mathbb{P}}g_{J,M}(C_{n,J}) \to a_{J,M}.$$

Together with (21), this yields

$$\lim_n \mathbb{E}_{\mathbb{Q}}\{L_n - g_{J,M}(C_{n,J})\}^2 = e^S - a_{J,M}.$$

First let $M\to\infty$, so that $a_{J,M}\uparrow\exp(\sum_{w\in J}\mu_w\delta_w^2)$, and then let J exhaust the words. The residual tends to zero. The finite approximants converge in distribution to $g_{J,M}(X_J)$, which in these successive limits converges in L^2 to L_∞. Approximation in L^2 therefore proves $L_n \Rightarrow L_\infty$ under $\mathbb{Q}$. This is the Poisson small-subgraph-conditioning argument of Janson (1995). Since the limit has mean one and is strictly positive almost surely, Le Cam's first lemma gives mutual contiguity.

C.1 A CONSISTENT SUPERCRITICAL CYCLE TEST

Set $\gamma_n = m/n$, $\Lambda_{F,n}^{\mathrm{cyc}} = d_F^2 s_F^4/\gamma_n$, and $\Lambda_n^{\mathrm{cyc}} = \Lambda_A + \Lambda_{F,n}^{\mathrm{cyc}}$. Let $\bar{A} = A - d_A/n$ and $\bar{F} = F - d_F/m$. For one word w, sum the product of centered entries over injective embeddings of its typed cycle and normalize by its exact null standard deviation to obtain $Z_{w,n}$. Define the signed graph amplitude and the two-step feature amplitude by

$$g_A = \sqrt{d_A}\,s_A, \qquad g_{F,n} = \frac{d_F s_F^2}{\sqrt{\gamma_n}}, \qquad g_A^2 = \Lambda_A, \quad g_{F,n}^2 = \Lambda_{F,n}^{\mathrm{cyc}}.$$

For one representative of each dihedral class, define

$$S_{\ell,n} = \sum_{|w|=\ell}\frac{g_A^{2r(w)}g_{F,n}^{2q(w)}}{h(w)} = \frac{(\Lambda_n^{\mathrm{cyc}})^\ell}{2\ell},$$

$$\vartheta_{w,n} = \frac{h(w)^{-1/2}g_A^{r(w)}g_{F,n}^{q(w)}}{\sqrt{S_{\ell,n}}}, \qquad T_{\ell,n} = \sum_w \vartheta_{w,n}Z_{w,n}.$$

The equality is the word-orbit identity for $\ell \geq 3$. Words containing a zero-signal channel have weight zero. Uniformly over the other words, the exact embedding count and Bernoulli null variances give

$$\mathbb{E}_{\mathbb{P}}Z_{w,n} = \big(1 + O_\theta(\ell^2/n)\big)\,\frac{g_A^{r(w)}g_{F,n}^{q(w)}}{\sqrt{h(w)}}. \tag{30}$$

Indeed, the count is $(n)_\ell(m)_{q(w)}/h(w)$, and the null variance per primitive edge is $(d_A/n)(1 - d_A/n)$ or $(d_F/m)(1-d_F/m)$.

Lemma 4 (Cycle-overlap excess)**.** *Let two injectively embedded typed cycles have macro-length ℓ and at least one common observed edge. If they are not the same embedded cycle up to a dihedral reparametrization, their quotient union H satisfies*

$$|E(H)| - |V(H)| \geq 1. \tag{31}$$

There are at most $(C\ell)^{C\ell}$ typed quotient records. After summing their embeddings and normalizing the two cycle polynomials by their null standard deviations, all such proper overlaps contribute $O((C\ell)^{C\ell}/n)$ to the absolute planted covariance. For identical embeddings, the only nonconstant label-even correction has magnitude at most ρ_1^ℓ for some fixed $\rho_1 \in (\max\{|s_A|, s_F^2\}, 1)$ and all large n.

Proof. Each expanded object is a simple cycle with at most 2ℓ vertices. Since the two cycles share an edge, their union is connected. Start with the first cycle, for which $|E| - |V| = 0$. If the second cycle is not identical, one of its maximal paths outside the first has two endpoints in the graph already exposed and all interior vertices new. Adding such an ear introduces one more edge than vertex; every further ear has nonnegative excess. This proves (31). Equality covers two cycles sharing a single path, including a path ending through either half of an F wedge. Reversal changes only the parametrization. If no outside ear exists, the edge sets coincide and simplicity makes the embeddings identical up to the excluded dihedral reparametrization.

A quotient record consists of the two words, a type-preserving equivalence relation on their at most 4ℓ expanded vertices, and the two cyclic orders and orientations. The crude bound $(C\ell)^{C\ell}$ therefore covers all records, including multiple common paths and half-wedge endpoints. If a record has v_I, v_F quotient vertices and e_A, e_F distinct observed edges, it has at most $n^{v_I} m^{v_F}$ embeddings. Conditional independence of distinct Bernoulli edges and

$$|\mathbb{E}[\bar{A}_e^r \mid \sigma]| \le C/n, \qquad |\mathbb{E}[\bar{F}_e^r \mid \sigma, \tau]| \le C/m, \qquad r \in \{1, 2\},$$

bound its moment by $(C/n)^{e_A}(C/m)^{e_F}$. Since m/n stays bounded above and below, the absolute normalized contribution of this fixed record is bounded by

$$C^\ell n^{v_I + v_F - e_A - e_F} \le \frac{C^\ell}{n}. \tag{32}$$

Here the two exact null standard deviations and the fixed word amplitudes are absorbed into C^ℓ; the record includes the quotient multiplicity. Combining (32) with the record count gives the claimed $O((C\ell)^{C\ell}/n)$ bound. Edge-disjoint embeddings, including pairs that share vertices, have zero covariance: conditional edge independence factorizes them, and the conditional mean of either cycle is deterministic because every latent sign appears twice.

For an identical expanded cycle the common-edge incidence matrix has a one-dimensional binary kernel. Its two label-even choices are the constant term and the full cycle. The latter contributes $\kappa_{A,n}^{r(w)} \kappa_{F,n}^{2q(w)}$. Since $\kappa_{A,n} \to s_A$, $\kappa_{F,n} \to s_F$, and $r(w) + q(w) = \ell$, its magnitude is at most ρ_1^ℓ uniformly over words for all large n. □

For $\ell \to \infty$ with $\ell \log \ell = o(\log n)$, injective embeddings give

$$\mathbb{E}_{\mathbb{Q}} T_{\ell,n} = 0, \quad \mathrm{Var}_{\mathbb{Q}}(T_{\ell,n}) = 1, \quad \mathbb{E}_{\mathbb{P}} T_{\ell,n} = (1 + o(1))\sqrt{S_{\ell,n}}.$$

For the planted covariance, Lemma 4 shows that identical embeddings contribute $1 + O(\rho_1^\ell) + o(1)$, while proper overlaps total $O((C\ell)^{C\ell}/n)$. Thus $\mathrm{Var}_{\mathbb{P}}(T_{\ell,n}) = 1 + o(1)$. One explicit uniform choice is

$$\ell_n = \max\left\{3, \left\lfloor \sqrt{\frac{\log(e+n)}{\log\log(e^e+n)}} \right\rfloor\right\}.$$

It satisfies $\ell_n \log \ell_n = o(\log n)$. For every fixed $\Lambda > 1$, eventually $\Lambda_n^{\mathrm{cyc}} \ge 1 + (\Lambda - 1)/2$, so $S_{\ell_n,n} \to \infty$. Rejecting the null when $T_{\ell_n,n} > \frac{1}{2}\sqrt{S_{\ell_n,n}}$ gives a consistent test by Chebyshev's inequality. The cycle weights use the observed ratio γ_n; the argument holds for every sequence $\gamma_n \to \gamma$.

Together with Section B, the subcritical cycle product proves mutual contiguity in Theorem 1(1). The supercritical cycle test proves asymptotic singularity in Theorem 1(2); the weak-recovery part of that statement is completed by the path construction in Section D.

D Self-avoiding paths and polynomial-time recovery

D.1 Typed paths and overlap diagrams

Assume first that both channels have nonzero signal. With $\Delta_A = d_A s_A$ and $\Delta_F = d_F s_F$, define unbiased centered edge scores

$$\widehat{A}_{ij} = \frac{n}{\Delta_A}\left(A_{ij} - \frac{d_A}{n}\right), \qquad \widehat{F}_{i\mu} = \frac{m}{\Delta_F}\left(F_{i\mu} - \frac{d_F}{m}\right).$$

Their conditional means are respectively $\sigma_i\sigma_j$ and $\sigma_i\tau_\mu$. Their conditional second moments are

$$\mathbb{E}[\widehat{A}_{ij}^2 \mid \sigma] = c_{A,n}\frac{n}{\Lambda_A}(1+\kappa_{A,n}\sigma_i\sigma_j),$$
$$\mathbb{E}[\widehat{F}_{i\mu}^2 \mid \sigma,\tau] = c_{F,n}\frac{m}{d_F s_F^2}(1+\kappa_{F,n}\sigma_i\tau_\mu),$$

where $c_{A,n}, c_{F,n} \to 1$ and $\kappa_{A,n} \to s_A$, $\kappa_{F,n} \to s_F$. Explicitly,

$$c_{A,n} = 1-\frac{d_A}{n}, \qquad \kappa_{A,n} = s_A\frac{1-2d_A/n}{1-d_A/n}, \qquad c_{F,n} = 1-\frac{d_F}{m}, \qquad \kappa_{F,n} = s_F\frac{1-2d_F/m}{1-d_F/m}. \tag{33}$$

For a macro-word $\omega \in \{A,F\}^k$, let $\mathcal{L}_\omega(i,j)$ be the self-avoiding paths from item i to item j in which an A letter is one direct edge and an F letter is a two-edge wedge through a distinct feature. If $q(\omega)$ is the number of wedges, put $r(\omega) = k - q(\omega)$ and let $(z)_h$ denote the falling factorial. Then

$$N_\omega = |\mathcal{L}_\omega(i,j)| = (n-2)_{k-1}(m)_{q(\omega)}.$$

Let p_π be the product of the edge scores along π. Absorb the finite-n second moments into

$$\Lambda_{A,n}^{\mathrm{path}} = \frac{\Lambda_A}{c_{A,n}}, \qquad \Lambda_{F,n}^{\mathrm{path}} = \frac{(d_F s_F^2)^2 n}{c_{F,n}^2 m}, \qquad \Lambda_n^{\mathrm{path}} = \Lambda_{A,n}^{\mathrm{path}} + \Lambda_{F,n}^{\mathrm{path}},$$

and set $\eta_{A,n} = \Lambda_{A,n}^{\mathrm{path}}/\Lambda_n^{\mathrm{path}}$ and $\eta_{F,n} = \Lambda_{F,n}^{\mathrm{path}}/\Lambda_n^{\mathrm{path}}$. The path polynomial is

$$\mathsf{P}_{ij}^{(k)} = \sum_\omega \frac{\eta_{A,n}^{r(\omega)}\eta_{F,n}^{q(\omega)}}{N_\omega} \sum_{\pi\in\mathcal{L}_\omega(i,j)} p_\pi. \tag{34}$$

Telescoping conditional means show exactly that $\mathbb{E}[\mathsf{P}_{ij}^{(k)} \mid \sigma] = \sigma_i\sigma_j$. Write $\eta_\omega = \eta_{A,n}^{r(\omega)}\eta_{F,n}^{q(\omega)}$ and use $\Lambda_{\chi,n}^{\mathrm{path}}$ for the corresponding finite-n signal, $\chi \in \{A,F\}$. If a signal channel vanishes, give it weight zero, omit its undefined edge score, and restrict all word sums to the active set $\mathcal{C}_{+,n} = \{\chi : \eta_{\chi,n} > 0\}$. Then $\sum_\omega \eta_\omega = 1$ in the two-channel and both one-channel cases.

We now define path overlaps independently of their embedding into $V \sqcup U$. For path copy $a \in \{1,2\}$ introduce item coordinates $I_0^{(a)}, \dots, I_k^{(a)}$. At macro position t, an A letter places the primitive edge $I_{t-1}^{(a)}I_t^{(a)}$; an F letter introduces a feature coordinate $U_t^{(a)}$ and the two primitive edges $I_{t-1}^{(a)}U_t^{(a)}$ and $U_t^{(a)}I_t^{(a)}$. The prescribed matches are $I_0^{(1)} \sim I_0^{(2)} = i$ and $I_k^{(1)} \sim I_k^{(2)} = j$.

A *valid overlap diagram* is a type-preserving partial matching between the remaining coordinates of the two abstract paths, with each coordinate used at most once. Equivalently, it is the quotient obtained by identifying the matched pairs. A primitive edge is common when its two endpoints are matched to the endpoints of an edge in the other path, in either orientation. Let $E_\cap$ be these common observed edges, E_0 the edges used once, and $G_\cap = (V(E_\cap), E_\cap)$. Write e_A, e_F for the numbers of direct and bipartite edges in $E_\cap$, $c_\cap$ for the number of its nonempty components, t_∂ for the number of i, j incident to $E_\cap$, and v_0 for the number of further shared vertices not incident to a common edge. Finally, let v_I, v_F count all shared nonendpoint item and feature vertices, including the v_0 isolated identifications. These coordinates will also make the reconstruction and multiplicity bounds below explicit.

D.2 Forest parity

Lemma 5 (Forest–parity reduction)**.** *For any fixed valid diagram of two paths with endpoints i, j, the constant label term is the unique surviving expansion. Consequently,*

$$\mathbb{E}[p_\pi p_{\pi'}] = \left(\frac{n}{\Lambda_{A,n}^{\mathrm{path}}}\right)^{e_A}\left(\frac{c_{F,n}m}{d_F s_F^2}\right)^{e_F}. \tag{35}$$

A zero-signal channel is absent from positive-weight diagrams.

Proof. Index the choice of the label-dependent term in the second moment of each common edge by $x \in \mathbb{F}_2^{E_\cap}$. Averaging the product over all latent signs retains exactly the choices for which every sign has even exponent, namely

$$B_\cap x = B_0 \mathbf{1} = 0 \quad \text{over } \mathbb{F}_2, \tag{36}$$

where $B_\cap, B_0$ are the typed vertex–edge incidence matrices. To see $B_0\mathbf{1} = 0$, add the edge indicators of the two paths over $\mathbb{F}_2$. Their common edges cancel, and their two identical boundary vectors $\mathbf{e}_i + \mathbf{e}_j$ also cancel. Thus the edges used once form an even-degree graph. The restriction of the quotient map to either simple path is injective, so $G_\cap$ is a forest of typed path components. Hence

$$\dim \ker B_\cap = |E_\cap| - |V(E_\cap)| + c_\cap = 0.$$

Equation (36) therefore has the unique solution $x = 0$. Multiplying the constant parts of the common-edge second moments in (33) proves (35), including reversed segments and one-half-wedge overlaps. □

D.3 Vertex loss and common components

Lemma 6 (Typed vertex loss)**.** *For a valid diagram,*

$$v_I + v_F = e_A + e_F + c_\cap - t_\partial + v_0. \tag{37}$$

The number of embedded path pairs realizing it is exactly

$$(n-2)_{2(k-1)-v_I} (m)_{q(\omega)+q(\omega')-v_F}, \tag{38}$$

and its ratio to $N_\omega N_{\omega'}$ is, uniformly for $k = O(\log n)$,

$$\left(1 + O\left(\frac{k^2}{n}\right)\right) n^{-v_I} m^{-v_F}. \tag{39}$$

If a common-edge component contains both i and j, the two expanded paths coincide fully. After excluding that class, every boundary component contains exactly one prescribed endpoint, and therefore

$$c_0 := c_\cap - t_\partial \geq 0 \tag{40}$$

is precisely the number of interior common-edge components.

Proof. The forest $G_\cap$ has $e_A + e_F + c_\cap$ incident vertices. Exactly t_∂ of them are prescribed endpoints; adding the v_0 isolated shared nonendpoint vertices gives (37) while retaining the item and feature counts separately. The quotient diagram has $2(k-1) - v_I$ free item classes and $q(\omega) + q(\omega') - v_F$ free feature classes. Injectively assigning these classes proves (38); the falling-factorial expansion $(N)_r = N^r \exp\{O(r^2/N)\}$, with $m \asymp n$, gives (39).

If one component contains both endpoints, it contains an i–j path. The unique i–j path in either simple expanded path is the entire path, so both paths coincide. Otherwise no component contains both endpoints. Each endpoint incident to $E_\cap$ then belongs to a distinct boundary component, which proves (40). Fully coincident paths ($c_\cap = 1$, $t_\partial = 2$) are evaluated separately. □

Lemma 7 (Maximal-component decomposition)**.** *Every nonempty component of $G_\cap$ is a maximal common subpath, traversed in the same or reverse direction by the two expanded paths. In macro coordinates it consists of complete common A steps and F wedges, except for at most one common F half-edge at each of its two ends. Hence the states "complete A," "complete F," left/right half-wedge, orientation, and anchor/isolated vertex exhaust all local overlap configurations.*

Proof. The common-edge graph is a subgraph of either expanded simple path, so every vertex has common degree at most two and each nonempty component is a path. Maximality makes its common edges consecutive in each copy; their orders either agree or are reversed. Every feature vertex internal to such a component has both incident edges common and therefore forms a complete F wedge. A feature vertex with only one common incident edge must be an endpoint of the component, producing exactly a left or right half-wedge. The two component endpoints therefore exhaust all partial wedges. Item vertices are macro boundaries, and shared vertices incident to no common edge are precisely the isolated records. □

D.4 CANONICAL RECORDS AND SKELETON RECONSTRUCTION

Lemma 8 (Canonical records reconstruct the diagram)**.** *For fixed words, every proper overlap diagram is determined by canonical records using at most Ck^4 choices for each interior common-edge component and Ck^2 choices for each isolated shared nonendpoint vertex. Conversely, each consistent collection of these records determines at most one valid overlap diagram. A boundary component has only a finite number of states once its number of complete macro-steps is specified.*

Proof. Index the expanded vertices and edges in their path order; an expanded path has at most $2k+1$ vertices. Order the interior common-edge components by the position of their first edge on the first path. For each component record its two endpoint positions on each path, its relative orientation, whether either endpoint cuts the left or right half of an F-wedge, and its boundary mask in $\{i, j\}$. The four positions give $O(k^4)$ choices, and orientation, half-wedge phases, and the boundary mask range over fixed finite sets. A boundary component is anchored at i or j; after its length is specified, only those finite orientation and phase states remain. Indeed, its intervals are prefixes of both paths at i, or suffixes at j; the length and endpoint phases fix their other endpoints, with no free offset. For an isolated shared vertex, its pair of expanded positions and its type give $O(k^2)$ choices.

By Lemma 7, each common component is a path. Once its two endpoint positions and relative orientation are fixed, consecutive primitive edges on the first path must be paired consecutively, in the recorded direction, with those on the second. Every internal feature then has common degree two and is a complete wedge; a common feature of degree one can occur only at a recorded F_{L} or F_{R} endpoint. Thus the record forces every paired vertex and edge incident to that component. The isolated records force all remaining shared vertices. Their union is the entire type-preserving equivalence relation, so a consistent collection has at most one preimage diagram.

Ordering components by their first edge on the first path removes permutation multiplicity. Inconsistent records—for example, those identifying an item with a feature, using one coordinate twice, or placing two maximal components through the same common edge—have no preimage. Counting them can only increase the upper bound. The phase and orientation fields explicitly cover one-half-wedge and reversed components. □

Lemma 9 (Skeleton correspondence)**.** *Erase only the A/F types of complete common macro-steps from a proper canonical record, while retaining their paired macro positions, endpoint phases, relative orientations, and all isolated shared-vertex records. Call the result a* skeleton $\mathcal{D}$*. Its compatible preimages are encoded by*

1. *one common active-channel type for every paired complete macro position;*
2. *one active-channel type for every unconstrained macro position of each path; and*
3. *at most $C^{c_\cap + v_0}$ local phase completions.*

After these choices, the two words and their expanded overlap diagram are determined. The independent macro-type assignments and $C^{c_\cap + v_0}$ local completions bound its multiplicity.

Proof. Use macro coordinates (p, ϕ), where $p \in \{1, \dots, k\}$ and $\phi \in \{I, F_{\mathrm{L}}, F_{\mathrm{R}}\}$ records an item boundary or a side of an F wedge. These coordinates remain fixed across expanded path lengths. Reconstruct a preimage in three steps:

1. Fix the recorded endpoint phases and orientations. Mark each macro slot touched by a partial wedge or an isolated shared feature as F. An isolated shared item marks an item boundary and imposes no letter.
2. Assign one common letter to each paired complete macro position, and one letter to each remaining unconstrained position in each path. Expand A into one item edge and F into an item–feature–item wedge. Macro positions do not change when a feature vertex is inserted.
3. Identify the vertices along each common component in its recorded orientation, then apply the isolated-vertex identifications. Discard a record if it violates types, simplicity, or the specified maximal components.

Each surviving assignment determines both words and the full vertex equivalence relation. An F_{L} or F_{R} endpoint selects the primitive half-edge even for a reversed component.

The type-constraining slots—complete common steps, common half-wedges, and isolated shared features—are disjoint across records. Two records using the same such slot would share its feature or an observed edge, and would belong to one common component; incompatible meetings are discarded. In contrast, an isolated shared item fixes a macro boundary, not the types of its incident steps. It may lie at the other end of a partially shared wedge without imposing another letter constraint or duplicating a word weight. The same distinction applies to item endpoints of common components. Consequently every residual ambiguity is a fixed phase/orientation choice attached to one component or isolated record. Multiplying those choices gives $C^{c_\cap + v_0}$, and the stated data reconstruct the vertex equivalence relation and hence the expanded diagram. □

A worked overlap diagram. Consider two paths of macro-word $FAFAFA$:

$$\begin{aligned} \pi &: i - \mu - a - b - \nu - c - d - \omega - e - j, \\ \pi' &: i - \mu - x - c - \nu - b - y - \omega - z - j. \end{aligned} \tag{41}$$

Greek symbols are features; all other symbols are items, with different symbols denoting different vertices. The common subgraph has the boundary half-wedge $i - \mu$, the reversed interior wedge $b - \nu - c$, and the isolated shared feature ω. Their canonical records are:

Shared part	Paired macro positions	Contribution before word weights
$i - \mu$	first F step, left half in both copies	$c_{F,n}/z_F$
$b - \nu - c$	third step, reversed	$n^{-1}(\Lambda^{\mathrm{path}}_{F,n})^{-1}$
ω	feature at fifth step in each copy	m^{-1}

Here $z_F = d_F s_F^2$, $c_0 = 1$, $v_0 = 1$, and $(v_I, v_F, e_A, e_F, c_\cap, t_\partial) = (2, 3, 0, 3, 2, 1)$. The vertex-loss identity reads $2 + 3 = 0 + 3 + 2 - 1 + 1$. Up to the falling-factorial correction, the normalized moment is

$$n^{-2} m^{-3} \left(\frac{c_{F,n} m}{z_F} \right)^3 = \frac{c_{F,n}^3}{z_F^3 n^2}.$$

This equals the product of the table's three contributions. The boundary half-wedge and isolated feature force their two incident word letters to be F. In the skeleton the complete third step can instead be A or F; summing its weighted moment gives $(\Lambda^{\mathrm{path}}_n)^{-1}$. Thus the interior component and isolated identification leave two vertex penalties, regardless of the shared segment length.

D.5 Weighted skeleton summation

Proof of Lemma 2. Choose one anchor vertex in each common-edge component. A boundary anchor is prescribed; an interior anchor is a shared item or feature vertex and costs respectively n^{-1} or $m^{-1} = O(n^{-1})$ in (39). Every isolated shared vertex costs the same. Exposing the rest of a component along its common edges then pairs each repeated-edge moment in Lemma 5 with the item and feature vertex losses in Lemma 6. This keeps the n and m factors separate until a complete macro-step has been formed. A complete common A step contributes $1/\Lambda^{\mathrm{path}}_{A,n}$; a complete common F wedge contributes $1/\Lambda^{\mathrm{path}}_{F,n}$, because its two repeated half-edge factors are paired with one new item and one new feature identification.

The local factors are listed below. Put $z_F = d_F s_F^2$ on an active feature channel. The last column is the product of the middle two columns; for complete steps and anchors it records the residual after extracting the indicated canonical factor.

Local state	Vertex loss	Repeated-edge moment	Residual factor
Complete A step	n^{-1}	$n/\Lambda_{A,n}^{\mathrm{path}}$	1 after $(\Lambda_{A,n}^{\mathrm{path}})^{-1}$
Complete F wedge	$(nm)^{-1}$	$(c_{F,n}m/z_F)^2$	1 after $(\Lambda_{F,n}^{\mathrm{path}})^{-1}$
$F_{\mathrm{L}}/F_{\mathrm{R}}$ half-wedge	m^{-1} or n^{-1}	$c_{F,n}m/z_F$	$c_{F,n}/z_F$ or $c_{F,n}(m/n)/z_F$
Isolated vertex (I/F)	n^{-1} or m^{-1}	1	1 or n/m after n^{-1}
Anchor (boundary/interior)	1 or n^{-1}, m^{-1}	1	no boundary loss; interior n^{-1} times 1 or n/m
Reversal	unchanged	unchanged	orientation flag only

Lemma 7 shows that every maximal common component consists of an anchor, complete macro-steps, and at most one half-wedge at each end; isolated identifications are covered by the isolated-vertex row. The table therefore localizes every factor by component endpoints. Since $m/n \to \gamma$ and the active-channel parameters are fixed, all residual entries are uniformly bounded in n.

Let $\mathcal{P} = \{I, F_{\mathrm{L}}, F_{\mathrm{R}}\}$ record whether a component end lies at an item or cuts the left or right half of a wedge. Thus every component has an end state $\xi \in \Xi := (\mathcal{P}^2)^2 \times \{+,-\}$: the two copies' phases at its two ends and their relative orientation. For $\xi \in \Xi$, let $K_{\xi,n}$ be the product of the forced word weights, repeated half-edge moments, and m/n conversions left after removing the anchor loss and all complete macro-step factors. Each state contains only finitely many factors among $\eta_{A,n}, \eta_{F,n}, c_{F,n}/(d_F s_F^2), m/n$ and their reciprocals on an active channel. Hence, for the fixed parameter tuple,

$$\max_{\xi\in\Xi} |K_{\xi,n}| \le H_\theta < \infty. \tag{42}$$

Aspect-ratio factors are absorbed here once per component, so their number is proportional to the number of components.

Form the skeleton $\mathcal{D}$ of Lemma 9. Let $\mathcal{U}_1, \mathcal{U}_2$ be the macro positions in the two paths not constrained by a complete common step, partial-wedge state, or shared feature position. If the complete common steps have types $\chi_1, \dots, \chi_R$, then (35) and (39) give the termwise bound

$$\begin{aligned}\mathcal{W}_{\omega,\omega'}(\mathcal{D}) \le \exp\left\{\frac{Ck^2}{n}\right\} n^{-c_0-v_0} \prod_{p\in\mathcal{U}_1} \eta_{\omega_p,n} \prod_{p\in\mathcal{U}_2} \eta_{\omega'_p,n} \\ \times \prod_{h=1}^{R} \frac{\eta_{\chi_h,n}^2}{\Lambda_{\chi_h,n}^{\mathrm{path}}} \prod_{\nu=1}^{c_\cap} K_{\xi_\nu,n} \prod_{\rho=1}^{v_0} K_{\zeta_\rho,n},\end{aligned} \tag{43}$$

where $K_{I,n} = 1$ and $K_{F,n} = n/m$ for an isolated item or feature identification. An isolated shared feature forces one F letter in each path; its additional word-weight factor $\eta_{F,n}^2 \le 1$ has been upper-bounded by one in $K_{F,n}$. These two slots belong to no other type-constraining record, so this discards each such weight only once. Forced partial-wedge weights are included in $K_{\xi_\nu,n}$. The preimage description in Lemma 9 is exhaustive: a complete paired slot receives one common active type; a half-wedge or isolated shared feature forces its incident type to be F; and every remaining macro-position is an unconstrained A/F letter. These choices determine the two words and the diagram. The bounded local phase completions contribute at most $C^{c_\cap+v_0}$, so no further multiplicity depends on a component's length or on k.

Every factor in (43) is nonnegative after taking absolute values. Bound the local residuals by H_θ first and enlarge the compatible assignments to the Cartesian product of all active letters at the disjoint slots. This can only increase the sum, even when some assignments fail the reconstruction checks. The remaining factors separate by slot: every unconstrained letter sums to one. At a complete common position, summing its erased type over the active channels gives exactly

$$\sum_{\chi\in\mathcal{C}_{+,n}} \frac{\eta_{\chi,n}^2}{\Lambda_{\chi,n}^{\mathrm{path}}} = \frac{1}{\Lambda_n^{\mathrm{path}}}. \tag{44}$$

Indeed $\eta_{\chi,n} = \Lambda_{\chi,n}^{\mathrm{path}}/\Lambda_n^{\mathrm{path}}$ on every active channel. Let r_ν be the number of complete common macro-steps in component ν and let $W(\mathcal{D})$ sum (43) over all compatible active-channel assignments.

Enlarging H_θ by the fixed local constant from Lemma 9, equations (42) and (44) yield

$$W(\mathcal{D}) \le \exp\left\{\frac{Ck^2}{n}\right\} n^{-c_0-v_0} H_\theta^{c_\cap+v_0} \prod_{\nu=1}^{c_\cap} (\Lambda_n^{\mathrm{path}})^{-r_\nu}. \tag{45}$$

Thus (45) preserves the separate item/feature count until the bounded ratio n/m is absorbed.

Put $\varepsilon_\theta = (\Lambda(\theta) - 1)/2 > 0$. For all large n, $\Lambda_n^{\mathrm{path}} \ge 1 + \varepsilon_\theta$. Consequently the complete-step lengths of each boundary component sum as a geometric series, and the at most two boundary components contribute $O_\theta(1)$. Each interior component retains one anchor loss $O(n^{-1})$, and each isolated identification retains another. Combining these losses with the record counts in Lemma 8 proves (15). □

D.6 Path-pair second moment

Combining the overlap estimates gives the second-moment bound for recovery.

Lemma 10 (Typed path-pair bound)**.** *Every valid overlap diagram belongs to exactly one of the following classes:*

1. *$E_\cap = \varnothing$; the paths may share isolated vertices and the total normalized contribution of this class is at most one;*
2. *$E_\cap \ne \varnothing$ and no common-edge component contains both prescribed endpoints; its maximal components consist of complete A steps, complete F wedges, and at most one left/right half-wedge at each end, with either orientation, plus isolated shared vertices of either type;*
3. *one common-edge component contains both endpoints; then the two expanded paths and their macro-words coincide fully.*

For every fixed parameter tuple with $\Lambda > 1$, the second class with c_0 interior components and v_0 isolated shared nonendpoint vertices contributes at most (15)*. Consequently there are constants C_θ, C_* such that, uniformly for $k = O_\theta(\log n)$,*

$$\sum_{\omega,\omega'} \frac{\eta_\omega \eta_{\omega'}}{N_\omega N_{\omega'}} \sum_{\pi \in \mathcal{L}_\omega(i,j),\, \pi' \in \mathcal{L}_{\omega'}(i,j)} |\mathbb{E}[p_\pi p_{\pi'}]| \le C_\theta + C_\theta k^{C_*} \left(\frac{1}{n} + \frac{n}{(\Lambda_n^{\mathrm{path}})^k}\right).$$

The same statement holds when one signal channel vanishes after deleting all words that use it.

Proof. The three classes are disjoint by definition and exhaustive according as $E_\cap$ is empty, nonempty with no endpoint-spanning component, or contains an endpoint-spanning component. Lemma 6 shows that the third case is full coincidence. In the second case, Lemma 7 exhausts the local states; Lemmas 8 and 9 encode each diagram with its complete multiplicity; and Lemma 2 gives (15) after summing all compatible macro types.

In the first class no observed edge is repeated. Conditional on all latent labels, distinct edge scores are independent and the first-moment telescoping identity gives

$$\mathbb{E}[p_\pi p_{\pi'} \mid \sigma, \tau] = (\sigma_i \sigma_j)^2 = 1.$$

This remains true when the paths share isolated latent vertices, because no primitive observation is shared. Averaging the labels and using that the word weights sum to one bounds the class by one.

For proper overlaps, sum Lemma 2 over $c_0, v_0 \ge 0$. Since $k = O(\log n)$,

$$\sum_{c_0,v_0 \ge 0} C_\theta \left(\frac{C_\theta k^4}{n}\right)^{c_0} \left(\frac{C_\theta k^2}{n}\right)^{v_0} \le C_\theta + \frac{C_\theta k^{C_*}}{n}. \tag{46}$$

Finally, a fully coincident pair has $\omega = \omega'$ and $\pi = \pi'$. Using Lemmas 5 and 6, its total contribution is at most

$$Cn \sum_\omega \eta_\omega^2 (\Lambda_{A,n}^{\mathrm{path}})^{-r(\omega)} (\Lambda_{F,n}^{\mathrm{path}})^{-q(\omega)} = Cn \left(\sum_{\chi \in \mathcal{C}_{+,n}} \frac{\eta_{\chi,n}^2}{\Lambda_{\chi,n}^{\mathrm{path}}}\right)^k = \frac{Cn}{(\Lambda_n^{\mathrm{path}})^k}. \tag{47}$$

Adding the three classes proves the claim. □

To choose the path depth, calculate $\Lambda_n^{\mathrm{path}}$. If it is at most one, use independent random signs; otherwise set

$$k_n = \min\left\{ \left\lfloor \frac{\min\{n-2, m\}}{2} \right\rfloor, \left\lceil \frac{4\log n}{\log \Lambda_n^{\mathrm{path}}} \right\rceil \right\}. \tag{48}$$

For every fixed $\Lambda > 1$, eventually $k_n = O_\theta(\log n)$ and $(\Lambda_n^{\mathrm{path}})^{k_n} \geq n^4$. Hence $k_n^{C_*}\{n^{-1} + n/(\Lambda_n^{\mathrm{path}})^{k_n}\} = o(1)$. Let P be the matrix with off-diagonal entries (34) and zero diagonal. The lemma gives

$$\mathbb{E}\|P\|_F^2 \leq C_0 n^2, \qquad \mathbb{E}\langle P, \sigma\sigma^\top\rangle = n(n-1).$$

D.7 Color-coded polynomial-time approximation

A word with q wedges uses $L_q = k + q + 1$ expanded vertices. Color $V \sqcup U$ independently and uniformly with L_q colors. For a path π of that size write

$$Z_c(\pi) = p_q^{-1}\mathbf{1}\{\pi \text{ is colorful under } c\}, \qquad p_q = \frac{L_q!}{L_q^{L_q}}.$$

For a fixed source item, the colorful path sum is computed by a subset dynamic program with state

$$D_i(x, S, t, h),$$

the total score of colorful typed walks ending at x, using color set S, having completed t macro-steps and visited h feature vertices. Its transitions append an A edge, the left half of an F wedge, or its right half, only when the new color is absent from S. Every counted walk is therefore an expanded self-avoiding path, and the state space and transitions cost $2^{O(L_q)}\,\mathrm{poly}(n, k)$.

Lemma 11 (Color-coding second moment)**.** *For each $q \in \{0, \ldots, k\}$ average*

$$T = \left\lceil (k+1)e^{2k+1} \right\rceil$$

independent colorings, independently across different q, and replace every q-wedge path in (34) *by its averaged multiplier $T^{-1}\sum_{r=1}^{T} Z_{c_{q,r}}(\pi)$. Denote the resulting matrix by $\widehat{P}$ and set $\widetilde{P} = (\widehat{P} + \widehat{P}^\top)/2$. For every fixed supercritical parameter tuple,*

$$\mathbb{E}_c[\widetilde{P} \mid A, F] = P, \tag{49}$$

$$\mathbb{E}_{\mathbb{P},c}\|\widetilde{P} - P\|_F^2 \leq C_\theta n^2. \tag{50}$$

Consequently $\widetilde{P}$ satisfies the three moment assumptions of Lemma 12 and is computable in randomized polynomial time.

Proof. For every fixed path, $\mathbb{E}_c Z_c(\pi) = 1$, which proves conditional unbiasedness before and after symmetrization. For any two paths of the same expanded size,

$$\mathbb{E}_c[Z_c(\pi)Z_c(\pi')] = \frac{\Pr_c(\pi, \pi' \text{ both colorful})}{p_q^2} \leq \frac{1}{p_q},$$

because the event in the numerator is contained in the event that π is colorful. Thus the absolute covariance of their centered multipliers is at most p_q^{-1}. If $E_{ij}^{[q]}$ denotes the color-coding error in the q-wedge part of entry (i, j), independence of the T repetitions and Lemma 10 give

$$\mathbb{E}_{\mathbb{P},c}|E_{ij}^{[q]}|^2 \leq \frac{p_q^{-1}}{T} \sum_{\substack{\omega,\omega':\\ q(\omega)=q(\omega')=q}} \frac{\eta_\omega \eta_{\omega'}}{N_\omega N_{\omega'}} \sum_{\pi,\pi'} |\mathbb{E}_{\mathbb{P}}[p_\pi p_{\pi'}]| \tag{51}$$

$$\leq \frac{C_\theta p_q^{-1}}{T}. \tag{52}$$

Restricting the nonnegative path-pair sum to fixed q can only reduce it. The elementary bound $L! \geq (L/e)^L$ gives $p_q^{-1} \leq e^{L_q} \leq e^{2k+1}$, and hence the last display is at most $C_\theta/(k+1)$.

Conditional on the observations, the centered errors for different q are independent and have mean zero. Their cross terms therefore vanish after averaging colorings. Summing (52) over q gives

a uniform C_θ bound per off-diagonal entry, and summing the entries proves (50). Because P is symmetric, passing from $\widehat{P}$ to $\widetilde{P} = (\widehat{P} + \widehat{P}^\top)/2$ preserves (49), preserves its inner product with $\sigma\sigma^\top$, and contracts both its error and Frobenius norm.

Finally,

$$\mathbb{E}\|\widetilde{P}\|_F^2 \le 2\mathbb{E}\|P\|_F^2 + 2\mathbb{E}\|\widetilde{P} - P\|_F^2 \le C_\theta n^2,$$

while (49) gives $\mathbb{E}\langle \widetilde{P}, \sigma\sigma^\top\rangle = n(n-1)$, and Cauchy–Schwarz gives

$$\mathbb{E}\langle \widetilde{P}, \sigma\sigma^\top\rangle^2 \le n^2\mathbb{E}\|\widetilde{P}\|_F^2 \le C_\theta n^4.$$

Since $L_q \le 2k+1$, all colorings and subset dynamic programs cost $2^{O(k)}\,\mathrm{poly}(n) = n^{O_\theta(1)}$. □

D.8 Matrix-to-label rounding

Lemma 12 (Matrix-to-label rounding)**.** *Let $\widetilde{P}$ be a symmetric random matrix computable from the observations and internal randomness. Suppose that, for constants $a, b, c > 0$ and all large n,*

$$\mathbb{E}[\langle \widetilde{P}, \sigma\sigma^\top\rangle] \ge an^2, \quad \mathbb{E}[\langle \widetilde{P}, \sigma\sigma^\top\rangle^2] \le bn^4, \quad \mathbb{E}\|\widetilde{P}\|_F^2 \le cn^2.$$

All expectations include the planted observation law and internal randomness. Then a dyadic-margin minimum-Frobenius-norm SDP followed by Gaussian sign rounding uses only $(\widetilde{P}, n)$ and internal randomness and returns $\widehat{\sigma} \in \{\pm 1\}^n$ satisfying

$$\liminf_{n\to\infty} \mathbb{E}\,\mathrm{ov}(\widehat{\sigma}, \sigma) > 0.$$

The rounding runs in randomized polynomial time when $\widetilde{P}$ has a polynomial-time representation.

Proof. Let $X = \langle \widetilde{P}, \sigma\sigma^\top\rangle$, $X_+ = \max\{X, 0\}$, and $R = \|\widetilde{P}\|_F$. Since $X^2 \le n^2R^2$, the assumed Frobenius bound controls its second moment. Moreover, $\mathbb{E}X_+ \ge \mathbb{E}X$ and $\mathbb{E}X_+^2 \le \mathbb{E}X^2$. Paley–Zygmund therefore gives

$$\mathbb{P}\left(X_+ \ge \frac{a}{2}n^2\right) \ge \frac{a^2}{4b} =: p_1.$$

Choose M so that $c/M^2 \le p_1/2$. Markov's inequality gives $\mathbb{P}(R > Mn) \le p_1/2$. Hence $\mathcal{E} := \{X \ge an^2/2,\ R \le Mn\}$ has probability at least $p_1/2$, and on this event

$$X \ge 2\delta_* Rn, \tag{53}$$

where $\delta_* = a/(4M) > 0$. The algorithm samples a dyadic margin; δ_* identifies a successful grid point in the analysis.

Independently draw $J \in \{1, 2, \ldots\}$ with $\Pr(J = j) = 2^{-j}$ and put $J_n = \lceil \log_2 n \rceil$. If $J > J_n$, output independent random signs. Otherwise set $\delta_J = 2^{-J}$ and minimize $\frac{1}{2}\|Y\|_F^2$ subject to

$$Y \succeq 0, \qquad \mathrm{diag}(Y) = \mathbf{1}, \qquad \langle \widetilde{P}, Y\rangle \ge \delta_J Rn. \tag{54}$$

If $R = 0$, or numerical optimization fails within its polynomial budget, output independent random signs. The precision guarantee below ensures success on the event used in the analysis. Fix the finite integer j_* such that $2^{-j_*} \le \delta_*$. For all large n, $j_* \le J_n$. On $\mathcal{E} \cap \{J = j_*\}$, the ground-truth correlation matrix is feasible with a factor-two margin. The projection optimality condition for the minimum-norm feasible point gives $\langle Y, \sigma\sigma^\top\rangle \ge \|Y\|_F^2$, while the constraint and Cauchy–Schwarz give $\|Y\|_F \ge \delta_{j_*} n$. Draw $g \sim N(0, Y)$ and output $\widehat{\sigma}_i = \mathrm{sign}(g_i)$. Put $D = \mathrm{diag}(\sigma)$ and $Z = DYD \succeq 0$. Since $\arcsin$ is odd, the Gaussian sign identity gives

$$\mathbb{E}_g\langle \widehat{\sigma}, \sigma\rangle^2 = \frac{2}{\pi}\mathbf{1}^\top \arcsin^\circ(Z)\mathbf{1}.$$

For every $r \ge 0$, the Schur product theorem gives $Z^{\circ(2r+1)} \succeq 0$ and therefore $\mathbf{1}^\top Z^{\circ(2r+1)}\mathbf{1} \ge 0$. The nonnegative odd-power expansion of $\arcsin$ now yields

$$\mathbf{1}^\top \arcsin^\circ(Z)\mathbf{1} \ge \mathbf{1}^\top Z\mathbf{1} = \langle Y, \sigma\sigma^\top\rangle \ge \delta_{j_*}^2 n^2.$$

Because J is independent of the data,

$$\Pr(\mathcal{E} \cap \{J = j_*\}) \ge (p_1/2)2^{-j_*},$$

and hence, for the positive parameter-dependent constant $c_{\mathrm{round}} := (p_1/\pi)2^{-3j_*}$,

$$\mathbb{E}\langle\widehat{\sigma},\sigma\rangle^2 \geq c_{\mathrm{round}} n^2.$$

Finally, $|\langle\widehat{\sigma},\sigma\rangle| \leq n$ implies $\mathbb{E}|\langle\widehat{\sigma},\sigma\rangle|/n \geq c_{\mathrm{round}}$, giving nonvanishing expected overlap (see also Hopkins & Steurer, 2017; Lu & Sen, 2023).

For finite precision, normalize the constraint by R and write $Q = \widetilde{P}/R$, $\delta = \delta_J$, and $S = \sigma\sigma^\top$. On the successful event, $Z = (1-\delta/4)S + (\delta/4)I$ has unit diagonal, minimum eigenvalue at least $\delta/4$, and $\langle Q, Z\rangle \geq 3\delta n/2$, since $\|Q\|_F = 1$ and $|\operatorname{tr} Q| \leq \sqrt{n}$. Thus the feasible region has inverse-polynomial interior margins. If Y_* is its minimum-norm point and a feasible Y has objective error at most ε, strong convexity gives $\|Y - Y_*\|_F \leq \sqrt{2\varepsilon}$ and hence

$$\langle Y, S\rangle \geq \delta^2 n^2 - n\sqrt{2\varepsilon} \geq \tfrac{1}{2}\delta^2 n^2 \quad \text{when } \varepsilon \leq \delta^4 n^2/8.$$

Because $\delta \geq (2n)^{-1}$, this requires only inverse-polynomial accuracy. Weak SDP optimization can use a slightly tightened correlation constraint; PSD projection and diagonal normalization then absorb inverse-polynomial feasibility errors. The interior margins above allow these corrections within the displayed objective tolerance. This preserves the rounding bound up to a constant factor. Arithmetic is performed to polynomial precision for the supplied fixed parameters; sampling and coordinatewise rounding also take polynomial time. □

The color-coded matrix $\widetilde{P}$ of Lemma 11 satisfies Lemma 12's moment assumptions. Together with the computable depth (48), this gives one uniform randomized procedure taking (n, m, θ) as input. Its polynomial exponent may depend on the fixed distance of Λ above one. Together with the supercritical test in Section C, this completes Theorem 1(2).

E Protocols for the numerical illustrations

All experiments use sparse adjacency representations and independently sampled Bernoulli edges from (1)–(2). Unless stated otherwise, $m = n$, $d_A = d_F = 3$, and inference receives the generating channel correlations. Latent labels are used to evaluate returned scores; algorithm settings and eigenvector selection use the observations and supplied parameters. BP summaries include every terminal iterate, including runs reaching an iteration cap. Method comparisons use absolute hard-label correlation. The phase figure uses the amplitude-sensitive soft overlap $|n^{-1}\sum_i \sigma_i \tanh(h_i^{\mathrm{BP}})|$, which records the decay of the BP field continuously.

E.1 BP equations, initialization, and reproducibility

The exact posterior is invariant under a global sign flip, so its single-vertex marginals are uniform. Random initialization can select an oriented BP solution with algorithmic magnetization $m_x^{\mathrm{BP}} = \tanh h_x^{\mathrm{BP}}$, evaluated up to a global sign.

For an observed edge xy, write $s_{xy} = s_A$ on the item graph and $s_{xy} = s_F$ on the incidence graph, and define $\phi_s(h) = \operatorname{atanh}\{s\tanh(h)\}$. If $h^{(t)}_{x\to y}$ is the half log-likelihood field sent from x to y, one undamped sparse binary cavity update is

$$\widetilde{h}^{(t+1)}_{x\to y} = \sum_{z\in\partial x\setminus\{y\}} \phi_{s_{xz}}(h^{(t)}_{z\to x}) - b^{(t)}_{\mathrm{type}(x)}. \tag{55}$$

With $\gamma_n = m/n$, the nonedge fields used in the experiments are

$$\widetilde{b}^{(t)}_I = d_A s_A \bar{m}^{(t)}_I + d_F s_F \bar{m}^{(t)}_U, \qquad \widetilde{b}^{(t)}_U = \frac{d_F s_F}{\gamma_n}\bar{m}^{(t)}_I, \tag{56}$$
$$\bar{m}^{(t)}_I = \frac{1}{n}\sum_{i\in V}\tanh H^{(t)}_i, \qquad \bar{m}^{(t)}_U = \frac{1}{m}\sum_{\mu\in U}\tanh H^{(t)}_\mu,$$

where $H^{(t)}_x = \sum_{z\in\partial x}\phi_{s_{xz}}(h^{(t)}_{z\to x}) - b^{(t-1)}_{\mathrm{type}(x)}$. The implemented update is $h^{(t+1)} = 0.55h^{(t)} + 0.45\widetilde{h}^{(t+1)}$ and $b^{(t)} = 0.55b^{(t-1)} + 0.45\widetilde{b}^{(t)}$, starting from $b^{(-1)} = 0$. Directed messages are

initialized by seeded pseudorandom draws from $\mathrm{Unif}[-0.02, 0.02]$ and then centered by subtracting their empirical mean. Iteration stops when the RMS message change is below 10^{-7} or after 1,400 iterations. Single-view BP removes the other edge type and its corresponding terms in (56).

All runs used CPU process parallelism and no GPU. The displayed phase plane and rate-ratio plane each contain 10,201 instances; the cuts contain 9,015 runs; the reported baseline suite contains 1,750 BP/fusion/projection evaluations on 250 sampled graphs and 100 additional BH evaluations on their joint-only subset; and the two controls in Figure 7 contain 605 runs each. The independent finite-size study in Section E.3.2 adds 1,400 graphs and 2,800 paired protocol outputs, requiring 2,081 BP calls after reusing 719 already-converged outputs.

E.2 Joint-value baselines

The main baseline experiment uses five channel compositions, 50 independent instances per composition, and $n = m = 5000$. The seven BP/fusion/projection methods use the same sampled edge lists at each composition/seed coordinate; BH is evaluated on the two joint-only compositions. Joint, graph-only, and feature-only BP use the sparse binary cavity update with at most 1400 iterations, damping 0.55, and root-mean-square (RMS) stopping tolerance 10^{-7}. Late fusion centers and normalizes the two single-view score vectors, aligns their unidentifiable global signs by the sign of their score inner product, and adds them.

Write $r_F = F\mathbf{1}$, $c_F = F^\top \mathbf{1}$, and $M_F = \mathbf{1}^\top r_F$. The zero-diagonal common-neighbor operator and its bipartite configuration-model correction are

$$P_0 = FF^\top - \mathrm{diag}(r_F),$$
$$q_F = \frac{\sum_\mu c_{F,\mu}(c_{F,\mu} - 1)}{M_F(M_F - 1)}, \qquad P_{\mathrm{dc}} = P_0 - q_F r_F r_F^\top + q_F \,\mathrm{diag}(r_F^2). \tag{57}$$

Naive projection uses the second algebraic eigenvector of P_0, excluding its Perron mode. Degree-corrected projection uses the leading eigenvector of P_{dc}. Using the supplied s_A, s_F, the combined projection baseline takes the leading eigenvector of

$$s_A\{A - \widehat{p}_A(\mathbf{1}\mathbf{1}^\top - I)\} + s_F^2 P_{\mathrm{dc}}, \qquad \widehat{p}_A = 2|E_A|/[n(n-1)].$$

Joint weighted Bethe–Hessian. Weighted BH is evaluated on the same 100 graphs in the two joint-only settings. Let s_{xy} be the weight of an observed edge. At radius $r = (1 + 10^{-4}) \max\{\sqrt{\rho(K)}, |s_A|, |s_F|\}$, define

$$H_s(r)_{xx} = 1 + \sum_{y \in \partial x} \frac{s_{xy}^2}{r^2 - s_{xy}^2}, \qquad H_s(r)_{xy} = -\frac{r s_{xy}}{r^2 - s_{xy}^2} \mathbf{1}\{xy \text{ observed}\}. \tag{58}$$

Both settings have positive channel correlations. We use the second smallest algebraic eigenpair to exclude the leading type-constant branch. Given its normalized vertex vector v, form $u_{x\to y} = (r v_y - s_{xy} v_x)/(r^2 - s_{xy}^2)$ and sum incoming $u_{y\to x}$ at each vertex. At readout, sums at vertices of total observed degree above 20 are set to zero; item labels are their signs, with zero mapped to $+1$. The eigensolver tolerance is 10^{-7}, its iteration cap is 5000, and all 100 runs returned eigenpairs with residual below 10^{-8}. The paired BH-minus-BP overlap differences are 0.0025 ± 0.0147 and 0.0173 ± 0.0175 (mean ± 2 SE).

Table 1 gives the two joint-only settings. The error bars use independent graph instances. The same experiment also includes graph-only and feature-only detectable controls: at $(\Lambda_A, \Lambda_F) = (1.20, 0.05)$, graph-only BP reaches 0.353 ± 0.026 and joint BP 0.416 ± 0.016; at $(0.05, 1.20)$, feature-only BP reaches 0.225 ± 0.029 and joint known-parameter BP 0.256 ± 0.032. At the jointly subcritical point $(0.40, 0.40)$, no method exceeds 0.065 mean overlap, versus at least 0.201 for joint BP in either joint-only setting. Known-parameter BP reaches the iteration cap on 10/50 balanced-rescue and 19/50 feature-rescue instances.

E.3 Phase and convergence diagnostics

Figure 3(a) uses nonlinear BP at $n = m = 200{,}000$ and $d_A = d_F = 3$. We sample a 101×101 grid in $x = \sqrt{\Lambda_A}$ and $y = \sqrt{\Lambda_F}$, then display it against $(\Lambda_A, \Lambda_F) = (x^2, y^2)$. Each cell is a

Table 1: Absolute item-label correlation, mean ± 2 standard errors (SE) over 50 independent instances. Both individual channels are subcritical in each column, which is indexed by (Λ_A, Λ_F).

Method	$(.60, .60)$	$(.20, .95)$
Joint BP (known parameters)	0.278 ± 0.033	0.201 ± 0.029
Joint weighted BH	0.281 ± 0.025	0.218 ± 0.023
Graph-only BP	0.036 ± 0.008	0.014 ± 0.003
Feature-only BP	0.030 ± 0.007	0.096 ± 0.018
Late fusion	0.036 ± 0.008	0.071 ± 0.014
Naive $FF^\top$	0.021 ± 0.004	0.035 ± 0.009
Degree-corrected $FF^\top$	0.023 ± 0.005	0.045 ± 0.009
$A+$ projected F	0.047 ± 0.010	0.062 ± 0.012

separately generated graph. Panel (b) independently samples a 101×101 grid in $\epsilon_A = p_{\text{out}}/p_{\text{in}}$ and $\epsilon_F = q_{\text{out}}/q_{\text{in}}$ at $d_A = 4, d_F = 2$ and $\gamma = 1$. The theoretical curve is $4s_A^2 + 4s_F^4 = 1$, with $s = (1-\epsilon)/(1+\epsilon)$. Panel (c) uses $\Lambda \in \{.700, .701, \ldots, 1.300\}$, five independent instances per coordinate, and $\Lambda_A/\Lambda \in \{.25, .50, .75\}$, for 9,015 runs in total. The contribution plane has 635/10,201 iteration-cap hits. The rate-ratio plane has 218/10,201, including 216 in $.9 \le \Lambda < 1.2$. The cuts have 1,435/9,015 (15.9%), ranging from $\Lambda = .927$ to 1.187. Graph size, initialization, and convergence affect the observed onset on this $.001$ grid.

E.3.1 Paired stopping-rule and initialization check

We reuse all five graphs at each of $\Lambda = .98, 1.02, 1.10$ from the equal-signal cut, keeping $n = m = 200{,}000$ and damping $.55$. Four settings give 60 evaluations: Original repeats the 1400-iteration, 10^{-7}-RMS protocol; Longer increases the cap to 5600 with the same initialization; Tighter uses 5600 iterations and tolerance 10^{-9}; Restart uses Longer with a second seeded initialization.

Table 2: BP sensitivity on paired phase-cut graphs. Soft overlap is mean ± 2 SE across five graphs.

Λ	Setting	Soft overlap	Cap hits
0.98	Original	0.00199 ± 0.00393	5/5
0.98	Longer	0.00334 ± 0.00664	0/5
0.98	Tighter	0.00333 ± 0.00665	2/5
0.98	Restart	0.00334 ± 0.00664	0/5
1.02	Original	0.01119 ± 0.01110	5/5
1.02	Longer	0.01225 ± 0.01030	3/5
1.02	Tighter	0.01225 ± 0.01030	3/5
1.02	Restart	0.01152 ± 0.01093	3/5
1.10	Original	0.07185 ± 0.00943	1/5
1.10	Longer	0.07185 ± 0.00943	0/5
1.10	Tighter	0.07185 ± 0.00943	0/5
1.10	Restart	0.07185 ± 0.00943	0/5

At $\Lambda = 1.10$, the mean soft overlap changes by less than 4×10^{-7} under all three alternatives. Nearer the boundary, Longer changes the means by $.00135$ at $.98$ and $.00106$ at 1.02. At 1.02, three of five runs reach the cap under each of Longer, Tighter, and Restart. Near the boundary, the observed overlaps remain sensitive to stopping rules and initialization.

E.3.2 Independent finite-size and paired-budget study

We sample 50 independent graphs at each of $n = m \in \{5{,}000, 20{,}000, 80{,}000, 200{,}000\}$ and $\Lambda \in \{.90, .95, .98, 1.00, 1.02, 1.05, 1.10\}$, giving 1,400 graphs. All samples are independent of the earlier phase cuts. The design fixes $d_A = d_F = 3$, positive channel correlations, and $\Lambda_A = \Lambda_F = \Lambda/2$. On each graph, we compare caps of 1,400 and 5,600 iterations with the same initialization, damping .55, and RMS tolerance 10^{-7}. If the shorter run already satisfies the stopping rule, the longer protocol returns the identical iterate, which we reuse. Otherwise, both calls start from the same seed. Records distinguish runs stopped by the cap from those meeting the RMS tolerance on the final permitted iteration.

Figure 6 reports both soft and hard overlap, cap frequencies, and the paired change in soft overlap. Each error bar is twice the standard error over 50 independent graphs at a fixed size and strength; the budget-change bars use the 50 within-graph differences. The computation uses 256 CPU worker processes with one serial BP kernel per worker.

Within this study, increasing the budget reduces cap hits from 681/1,400 (48.6%) to 361/1,400 (25.8%). At $n = 200{,}000$ and $\Lambda = 1.02$, cap hits fall from 41/50 to 20/50 and the paired soft-overlap change is $.000465 \pm .000603$. At the same size and $\Lambda = 1.10$, all longer runs converge and the paired change is $(-2.34 \pm 4.00) \times 10^{-7}$. Convergence also depends on size: at $\Lambda = 1.10$, the longer-budget cap counts are 30, 20, 4, and 0 out of 50 as n increases.

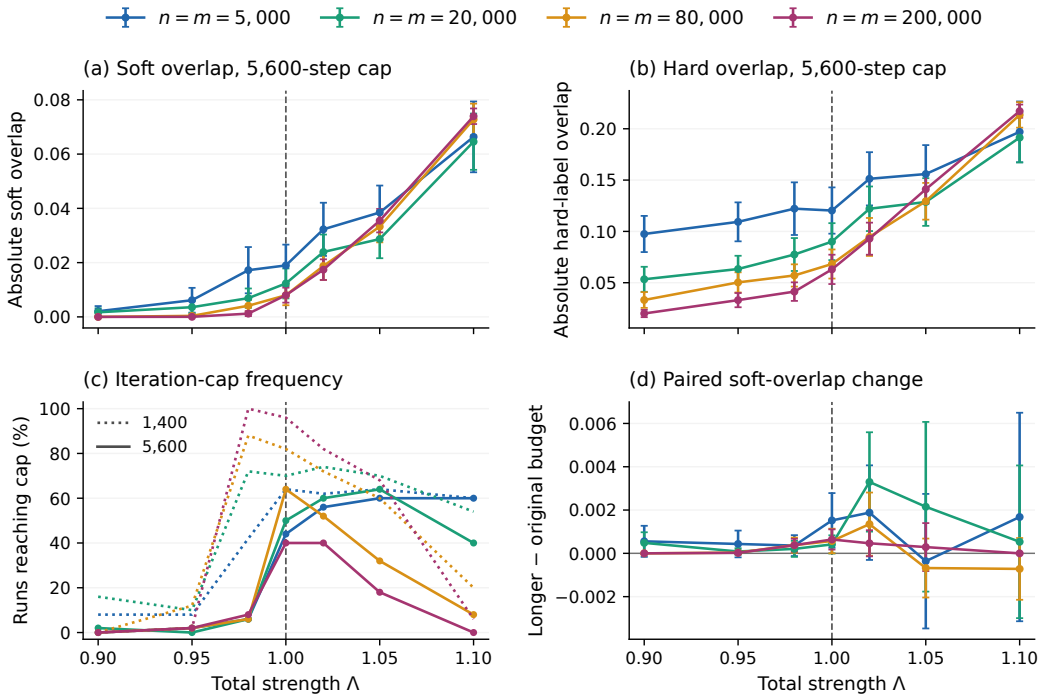

Figure 6: Finite-size BP near the additive boundary, with 50 independent graphs at each size and strength and $\Lambda_A = \Lambda_F = \Lambda/2$. **(a,b)** Soft and hard overlap under the 5,600-iteration cap. **(c)** Cap-hit frequency under the two budgets. **(d)** Paired soft-overlap difference, longer minus original budget. Bars show mean ± 2 SE. The vertical dashed line marks $\Lambda = 1$.

E.4 Aspect-ratio and signed-channel controls

Figure 7 varies the aspect ratio and the sign of the graph channel. Both controls use $n = 200{,}000$, $d_A = d_F = 3$, an even signal split $\Lambda_A = \Lambda_F = \Lambda/2$, and $\Lambda \in \{0.700, 0.705, \ldots, 1.300\}$, with five independent instances per point. The first sets $\gamma = 2$ (hence $m = 400{,}000$) and $s_A > 0$. The second sets $\gamma = 1$, $s_A < 0$, and $s_F > 0$; swapping the within- and between-item rates changes the sign

of s_A while preserving $\Lambda_A = d_A s_A^2$. The reference curve is the $\gamma = 1, s_A > 0$ cut from Figure 3, evaluated on its finer 0.001 grid.

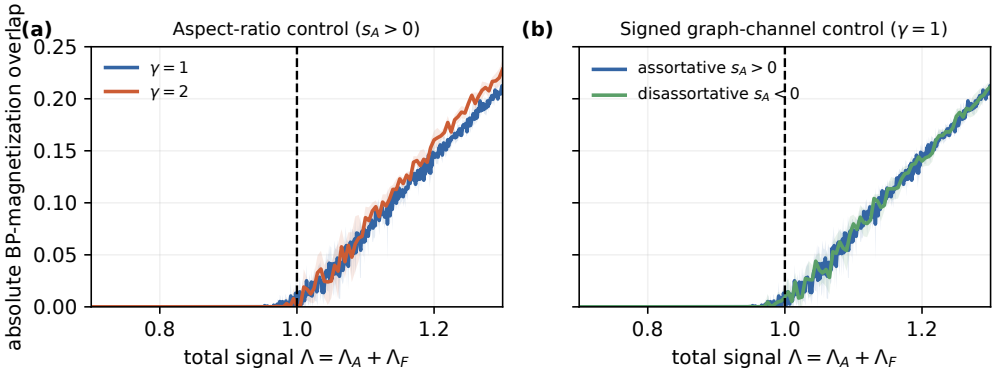

Figure 7: Additional nonlinear-BP phase cuts at $n = 200{,}000$ and $\Lambda_A = \Lambda_F = \Lambda/2$; curves show mean absolute BP-magnetization overlap and bands show ± 2 standard errors. **(a)** Changing the aspect ratio from $\gamma = 1$ to $\gamma = 2$ gives an empirical rise near the same predicted boundary. **(b)** Reversing the graph-channel sign gives a similar empirical onset, consistent with the threshold's dependence on s_A^2. The dashed line is the theoretical boundary.