WHAT STRUCTURE BUYS IN NETWORK CHANGE-POINT LOCALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Structural summaries can reduce the stochastic complexity of network change-point localization, but they can also erase the change. We study this tradeoff for one change in independent Bernoulli networks. Over unrestricted networks we prove the two-term minimax lower bound $\min\{T,\ \rho/\delta^2 + D\rho^2/(T\delta^4)\}$; for aligned disjoint aggregations the same lower bound holds with the number of structural cells in place of D. We see that valid side information can reduce an intrinsic dimension-dependent localization cost. A finite-sample analysis of the closely related Wang–Shao distinct-index contrast gives matching upper rates up to logarithms for aligned aggregation classes, while a deterministic misclustering lemma quantifies signal loss from an estimated partition. We also give results for general fixed linear maps, bounded frozen nonlinear features, and finite-library validity under an order-preserving parity split. Our experiments exhibit both sides of the retained-signal tradeoff, operationalize learned and selected structure, validate canonical raw-input baselines at strong signal, explain null-like boundary concentration of uninformative representations, and show calendar-robust late-September localization in Enron.

1 INTRODUCTION

Dynamic networks are routinely compressed before change-point analysis: edges may be kept individually, pooled by known groups, or summarized by degrees or learned features. Compression removes stochastic complexity, but only helps if the representation preserves the change. We ask when a supplied representation can improve localization of a single change in an independent Bernoulli network sequence.

If $D = \binom{n}{2}$, T is the number of snapshots, ρ bounds edge probabilities, and δ is the Frobenius jump, our unrestricted minimax benchmark is

$$\min\left\{T,\ \frac{\rho}{\delta^2} + \frac{D\rho^2}{T\delta^4}\right\}. \tag{1}$$

The first term is scalar resolution; the second is the price of an unknown high-dimensional change direction. With a known block partition, we show that D is replaced by the number of nonempty block pairs, so structural information can alter the localization rate itself.

Bernoulli networks add a complication: changing the mean also changes the variance, so the population expectation of a raw squared CUSUM can drift away from the true split. We use a split-dependent distinct-index correction closely related to the break-dating statistic of Wang & Shao (2023). For a linear map F, our rate is governed by the retained signal $\kappa_F^2 = \|F\Delta\|^2$, representation rank d_F, and, for general maps, edge leverage, so the central quantity is retained signal relative to the stochastic complexity that the representation removes.

Contributions. Our theoretical core is Theorems 6–7: the first makes the D-dependent localization term unavoidable over unrestricted Bernoulli changes; the second replaces D by the number of aligned structural cells, and Corollary 8 shows that normalized cell sums attain this improvement up to logarithms. Lemma 9 quantifies how much block signal survives an estimated partition. We also give a fixed-map upper bound with the exact range quantity $m_F(\Delta) = \|F^\top F\Delta\|_\infty$, a conservative effective-rank transfer bound for bounded frozen nonlinear features, and an order-preserving parity

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.

construction for honest finite-library selection. Empirically, we find that exact retained-signal sweeps show both harmful and beneficial compression; independent-pilot and parity-split experiments operationalize data-dependent structure; raw-input package checks remove the earlier sketching confound; and Enron calendar perturbations test the real-data feature.

Related work and scope. High-dimensional localization and sparse-projection work includes Wang & Samworth (2018); Kaul et al. (2021); Pilliat et al. (2023); Verzelen et al. (2023) give optimal univariate detection and localization; Wang & Shao (2023) derive the asymptotic $\|\Sigma\|_F^2/(T\|\Delta\|^4)$ break-dating scale corresponding to our second upper-bound term. Dimension-dependent terms also appear in detection boundaries (Liu et al., 2021); Yu (2020) reviews minimax rates, Cho & Li (2025) studies related detection lower bounds in a different model, and common-break models (Bai, 2010) lack the automatic mean–variance coupling of Bernoulli networks. Dynamic-network change-point methods span block, matrix, spectral, graph/kernel, and learned comparisons (Chen & Zhang, 2015; Bhattacharjee et al., 2020; Wang et al., 2021; Madrid Padilla et al., 2022; Huang et al., 2020; Sulem et al., 2024; Kei et al., 2025). Structural comparisons (Peel & Clauset, 2015) and sequence-of-graph methods (Zambon et al., 2019) are further comparators; these methods answer different questions about model estimation, graph geometry, or learned similarity. Our simpler model isolates the information-theoretic effect of a supplied representation; we give expanded positioning and interpretation in Appendix F.

2 Model and Corrected Contrast

We observe simple undirected graphs on $n \geq 3$ vertices with no self-loops and stack upper triangles as $Y_t \in \{0,1\}^D$. Coordinates are independent over time and edges with

$$\mathbb{E}Y_t = p_1 \mathbf{1}\{t \leq \tau\} + p_2 \mathbf{1}\{t > \tau\}, \qquad 0 \leq p_{j,e} \leq \rho \leq \tfrac{1}{2},$$

where, for a fixed $\eta \in (0, 1/2)$, $\tau \in \mathcal{K}_\eta = \{k : \eta T \leq k \leq (1-\eta)T,\ 2 \leq k \leq T-2\}$. The change location is the last time in the first regime, and all localization statements are relative to the scan over $\mathcal{K}_\eta$. Put $\Delta = p_2 - p_1$, $\delta^2 = 2\|\Delta\|^2$. For $F : \mathbb{R}^D \to \mathbb{R}^q$ with $\|F\|_{\mathrm{op}} \leq 1$, write $X_t = FY_t$, $H = F^\top F$, $d = \mathrm{rank}(F)$, and $\kappa^2 = \|F\Delta\|^2 > 0$.

For $r = T - k$, let $Q_k = kr\|\bar{X}_{1:k} - \bar{X}_{k+1:T}\|^2/T$. If $v_j = \mathrm{tr}\,\mathrm{Var}(X_t)$ in regime j, then for $k > \tau$,

$$\mathbb{E}Q_k - \mathbb{E}Q_\tau = \frac{k-\tau}{k}\{v_2 - v_1 - \tau\kappa^2\}, \tag{2}$$

with the time-reversed analogue for $k < \tau$. We see that variance drift can move the raw population maximizer away from τ (Appendix A). The drift is bilateral: the raw objective can rise to the right of τ when $v_2 - v_1 > \tau\kappa^2$ and to the left when $v_1 - v_2 > (T-\tau)\kappa^2$, which can cause asymptotic inconsistency (Proposition 14; Figures 6–7). We instead use

$$U_k = \frac{kr}{T}\left\{\frac{\sum_{s\neq t\leq k}\langle X_s, X_t\rangle}{k(k-1)} + \frac{\sum_{k<s\neq t\leq T}\langle X_s, X_t\rangle}{r(r-1)} - \frac{2\sum_{s\leq k<t}\langle X_s, X_t\rangle}{kr}\right\}, \tag{3}$$

and $\widehat{\tau}_U = \arg\max_{k\in\mathcal{K}_\eta} U_k$ (earliest maximizer). Within-block pairs are ordered, and equivalently $U_k = Q_k - (r/T)s_L^2 - (k/T)s_R^2$, where $s_L^2 = (k-1)^{-1}\sum_{t\leq k}\|X_t - \bar{X}_{1:k}\|^2$ and $s_R^2 = (r-1)^{-1}\sum_{t>k}\|X_t - \bar{X}_{k+1:T}\|^2$ are the unbiased within-block trace sample variances. Our estimator requires no estimate of ρ and no threshold when one change is assumed. Once features are formed, prefix sums give an $O(Tq)$ scan.

Proposition 1 (Population separation)**.** *For independent observations with the two mean levels above and arbitrary finite covariances, if Q_k^0 is the squared CUSUM of the means, then for $k > \tau$,*

$$\mathbb{E}U_k = Q_k^0 - \frac{\tau(k-\tau)(T-k)}{Tk(k-1)}\kappa^2, \qquad \mathbb{E}U_\tau = Q_\tau^0,$$

with the complementary expression by time reversal. Hence $\mathbb{E}(U_\tau - U_k) \geq c_\eta|k-\tau|\kappa^2$.

Our correction is unbiased at a pure split and has zero expectation under a constant mean; at a mixed split its expectation lies below the ordinary population CUSUM energy, which improves separation. In the notation of Wang & Shao (2023), $G_T(k)/T = U_k(k-1)(r-1)/(kr)$. We include this close predecessor in the experiments; the distinct-index construction is also closely connected to the two-sample statistic of Chen & Qin (2010).

3 FINITE-SAMPLE LOCALIZATION AND MINIMAX DIFFICULTY

Let $\ell = \max_e H_{ee}$ and $m_F(\Delta) = \|H\Delta\|_\infty$. The leverage ℓ, the largest squared feature loading of a single edge, remains small for many dense orthogonal embeddings, unlike an absolute row-sum bound. The elementary bound

$$m_F(\Delta) \le \sqrt{\ell}\,\kappa \tag{4}$$

gives a worst-case leverage simplification, but the exact Bernstein range is $m_F(\Delta)$.

Theorem 2 (Fixed-map upper bound with exact range term)**.** *For every deterministic F above and $x \ge 1$, with probability at least $1 - CTe^{-x}$,*

$$|\widehat{\tau}_U - \tau| \le C_\eta \left\{ \frac{\rho x + m_F(\Delta)x + \sqrt{\rho\ell/T}\, x^{3/2} + \ell x^2/T}{\kappa^2} + \frac{\rho^2(dx + x^2)}{T\kappa^4} \right\}. \tag{5}$$

If $T\rho \ge \ell x$, Theorem 2 also gives

$$|\widehat{\tau}_U - \tau| \le C_\eta \left\{ \frac{\rho x + m_F(\Delta)x}{\kappa^2} + \frac{\rho^2(dx + x^2)}{T\kappa^4} \right\}. \tag{6}$$

Substituting (4), we obtain the worst-case term $\sqrt{\ell}x/\kappa$, which above unit resolution can matter only when $\rho < \ell x$ (Appendix B). For averaging maps, $m_F(\Delta) \le C_0\rho$ directly.

Corollary 3 (Two-term bound for averaging maps)**.** *If additionally $m_F(\Delta) \le C_0\rho$ and $T\rho \ge \ell x$, then*

$$|\widehat{\tau}_U - \tau| \le C_{\eta,C_0} \left\{ \frac{\rho x}{\kappa^2} + \frac{\rho^2(dx + x^2)}{T\kappa^4} \right\} \tag{7}$$

with the same probability. The bound implies exact localization if its right side is less than one.

For raw edges, normalized degrees, and a disjoint aggregation with smallest cell size $N_{\min}$, the aggregation conditions are respectively $T\rho \ge x$, $(n-1)T\rho \ge x$, and $N_{\min}T\rho \ge x$. We see that d controls quadratic direction noise, whereas ℓ controls how strongly a single edge can enter the representation.

Corollary 4 (Expected-risk form)**.** *With $x = c_0 \log(eT)$ in Corollary 3 for a sufficiently large universal c_0,*

$$\mathbb{E}|\widehat{\tau}_U - \tau| \le C_{\eta,C_0} \min\left\{ T,\ \frac{\rho \log(eT)}{\kappa^2} + \frac{\rho^2\{d\log(eT) + \log^2(eT)\}}{T\kappa^4} \right\} + T^{-1}. \tag{8}$$

Corollary 5 (Independent-pilot representation)**.** *If $\widehat{F}$ is constructed from auxiliary data independent of $Y_{1:T}$, with $\|\widehat{F}\|_{\mathrm{op}} \le 1$ and $\widehat{\kappa} > 0$ given that data, then conditional on that data Theorem 2 holds with realized $(\widehat{d}, \widehat{\ell}, \widehat{\kappa})$; Corollary 3 holds whenever $\|\widehat{H}\Delta\|_\infty \le C_0\rho$ and $T\rho \ge \widehat{\ell}x$.*

Proof mechanism. For $u = |k - \tau|$ and $B = A_k - A_\tau$ (Lemma 16), $U_k - U_\tau$ splits into a linear term and a centered quadratic form in $B \otimes H$, whose deviations $\kappa\sqrt{\rho u x}$ and $\rho(\sqrt{dx} + x)\sqrt{u/T}$ we balance against the population loss $c_\eta u\kappa^2$ of Proposition 1; the bounded part $m_F(\Delta)x$ of the linear Bernstein term is at most ρx for averaging maps (Appendix B). Our second term is the finite-sample, heteroscedastic analogue of the Wang–Shao error scale (Appendix F.3), and Theorem 6 shows it is necessary.

Theorem 6 (Two-term Bernoulli lower bound)**.** *Let $0 < \rho \le 1/2$ and let $\mathcal{Q}_D(\rho, \delta)$ contain one-change Bernoulli sequences with $\tau \in \mathcal{K}_\eta$, probabilities in $[\rho/2, \rho]$, and exact jump δ. If $0 < \delta \le \rho\sqrt{2D}/4$ and $R_D = \min\{T,\ \rho/\delta^2 + D\rho^2/(T\delta^4)\} \ge C_\eta$, then*

$$\inf_{\widehat{\tau}} \sup_{Q \in \mathcal{Q}_D} \mathbb{E}_Q|\widehat{\tau} - \tau| \ge c_\eta R_D. \tag{9}$$

The dimension term is active heuristically over $\rho\sqrt{D} \lesssim T\delta^2 \lesssim \rho D$. Below the left boundary the risk saturates at order T; above the right boundary the scalar term ρ/δ^2 dominates. For raw edges $\kappa^2 = \delta^2/2$, so Corollary 4 and Theorem 6 match up to logarithms in their common feasible range. Our localization-specific lower-bound step conditions on the common suffix of two nearby change locations so uncertainty about random jump signs remains in the disputed block; we give full details in Appendix C.

Table 1: Canonical fixed maps (explicit forms in Appendix D). G is the unsigned vertex–edge incidence matrix; N_b is a block-pair edge count.

Map	$(d,\ \ell)$	Retained κ^2	Signal behavior	Consequence
Edges	$(D,\ 1)$	$\delta^2/2$	preserves every jump	robust reference
Count	$(1,\ 1/D)$	$(\mathbf{1}^\top\Delta)^2/D$	signed global mass	can erase balanced changes
Degrees	$(n,\ 1/(n-1))$	$\|(P_2-P_1)\mathbf{1}\|^2/[2(n-1)]$	balanced Chung–Lu scale	gain if degree-aligned
Blocks	$(d_{\mathrm{b}},\ 1/N_{\min})$	$\delta^2/2$ on block model	block-constant jump	sharp aligned gain

Theorem 7 (Aligned disjoint-aggregation lower bound)**.** *Partition the D edges into d nonempty cells of sizes N_b and $N_{\min} = \min_b N_b$, and restrict both regime means to be cell-constant. If $0 < \delta \le \rho\sqrt{2dN_{\min}}/4$ and $R_d = \min\{T,\ \rho/\delta^2 + d\rho^2/(T\delta^4)\} \ge C_\eta$, then $\inf_{\widehat{\tau}} \sup_Q \mathbb{E}_Q|\widehat{\tau} - \tau| \ge c_\eta R_d$. The normalized cell-sum map preserves every aligned jump and attains these terms up to logarithms when $N_{\min}T\rho \gtrsim \log(eT)$.*

Our results yield a clear criterion: structure is minimax-useful when it removes unresolved directions without erasing the jump. Theorem 7 makes the representation story two-sided for supplied disjoint aggregations: the number d of unresolved structural directions appears in both achievable and unavoidable error. We give the full fixed-map phase diagram and interpretation in Appendix F.

4 What structural information provides

For the nonnegative averaging maps below, $m_F(\Delta) \le \rho$. Each map in Table 1 has $\|F\|_{\mathrm{op}} \le 1$ and a nonnegative H with row sums at most one, so Corollary 3 applies directly (constants in Appendix D).

Corollary 8 (Known-partition minimax rate)**.** *For a fixed vertex partition with d_{b} nonempty block pairs and smallest pair size $N_{\min}$, if regime means are constant on each pair, Theorem 7 applies with $d = d_{\mathrm{b}}$ and Block U attains the same two terms up to logarithms whenever $N_{\min}T\rho \gtrsim \log(eT)$.*

Our result uses externally supplied labels; severe block imbalance narrows the admissible jump range through $N_{\min}$, and a misspecified partition enters through the retained signal κ_F, which Lemma 9 quantifies.

Lemma 9 (Retained signal under pilot-estimated partition)**.** *If the true jump is block-constant and an estimated partition misclassifies at most γn vertices after relabeling, then for the normalized block-pair map*

$$a_{F_{\widehat{z}}} = \frac{\|F_{\widehat{z}}\Delta\|^2}{\|\Delta\|^2} \ge 1 - 8\gamma\frac{n}{n-1}\frac{\Delta_{\max}^2}{\Delta_{\mathrm{rms}}^2}. \tag{10}$$

Here $\widehat{z}$ denotes the estimated labels, $\Delta_{\max} = \max_{a,b}|\Theta_{ab}|$ for the true block-jump matrix Θ, and $\Delta_{\mathrm{rms}}^2 = D^{-1}\|\Delta\|^2$.

Since the bound is deterministic, we combine it with Corollary 5 to convert any independent-pilot community-recovery guarantee into a localization guarantee without using the analysis sequence twice.

Proposition 10 (Balanced Chung–Lu signal)**.** *For $P(\theta)_{ij} = \theta_i\theta_j$ off-diagonal with $a\sqrt{\rho} \le \theta_i, \beta_i \le b\sqrt{\rho}$ for fixed $0 < a \le b < \infty$ and valid edge probabilities bounded by ρ, $\|P(\theta) - P(\beta)\|_{\mathrm{F}} \asymp n^{-1/2}\|\{P(\theta) - P(\beta)\}\mathbf{1}\|$ for large n. Hence $\kappa_{\deg} \asymp \delta$ and degrees inherit the n-dimensional upper rate.*

Retained signal versus removed complexity. With $s_0 = \|\Delta\|^2$ and $a_F = \|F\Delta\|^2/s_0$, Corollary 3 behaves like the unit-constant proxy $R_F(x) = \rho x/(a_F s_0) + \rho^2 x(d_F + x)/(Ta_F^2 s_0^2)$ of (18), with $(a_F, d_F) = (1, D)$ for raw edges. We see that compression shrinks the quadratic summand when $a_F > \sqrt{(d_F + x)/(D + x)}$, but every lossy map pays $1/a_F$ on the scalar summand: aligned maps with $a_F \approx 1$, such as known balanced blocks, remove the direction-search penalty at no scalar cost (minimax-sharply, by Corollary 8); maps with $0 < a_F < 1$ trade the two summands through the sufficient crossover (19); and $a_F = 0$ leaves the map blind. Figures 2–3 mark the proxy's unit-constant crossings (Appendix F.1).

5 FROZEN NONLINEAR REPRESENTATIONS AND HONEST DATA DEPENDENCE

We can also analyze a frozen nonlinear representation, at the price of its own tail behavior; below, $\mu_j = \mathbb{E}X_t$ and $\Sigma_j = \mathrm{Cov}(X_t)$ in regime j. For a predetermined $X_t = \psi(Y_t)$, write $\kappa_\psi = \|\mu_2 - \mu_1\|$, $s_\psi = \max_j \|\Sigma_j\|_{\mathrm{op}}$, $v_\psi = \max_j \|\Sigma_j\|_F$, and $r_2(\psi) = v_\psi^2/s_\psi^2$.

Theorem 11 (Bounded frozen-representation localization)**.** *Assume the X_t are independent over time with these two means, and apply (3) to X_t. If $\|X_t - \mu_{j(t)}\| \le B_\psi$ a.s., then for every $x \ge 1$, with probability at least $1 - CTe^{-x}$,*

$$|\widehat{\tau}_\psi - \tau| \le C_\eta \left\{ \frac{s_\psi x}{\kappa_\psi^2} + \frac{B_\psi x}{\kappa_\psi} + \frac{v_\psi^2 x + s_\psi^2 x^2}{T\kappa_\psi^4} + \frac{B_\psi \sqrt{s_\psi/T}\, x^{3/2} + B_\psi^2 x^2/T}{\kappa_\psi^2} \right\}. \tag{11}$$

The first and third terms have the same two-scale interpretation as the linear theory. The quadratic term depends on covariance effective rank through $v_\psi^2 = s_\psi^2 r_2(\psi)$; the B_ψ terms make this a conservative transfer result. For normalized or clipped frozen embeddings the B_ψ terms can be controlled directly; for raw edge features they are too crude, which is why we prove Theorem 2 with a sharper edge-level argument. Beyond kernel localization theory (Garreau & Arlot, 2018), our addition is the explicit retained-mean/effective-rank decomposition for a frozen network representation.

Order-preserving splitting. Let $\mathcal{E} = \{2, 4, \ldots\}$ and $\mathcal{O} = \{1, 3, \ldots\}$ index the even and odd snapshots. By independence the two subsequences are independent, each retains the ordering of the change, and the odd change location is $\tau_{\mathcal{O}} = \lceil \tau/2 \rceil$.

Proposition 12 (Order-preserving split transfer)**.** *Let $\widehat{\psi}$ be measurable with respect to the even snapshots only (it may be chosen from a finite library or trained by an arbitrary algorithm) and satisfy the applicable fixed-representation theorem on the odd subsequence. Assume $\eta T \ge 4$, run the odd scan with trimming $\eta/2$, and let $\widehat{\tau}_{\mathcal{O}}$ maximize the corrected contrast there. If that theorem gives odd-scale radius $R_{\widehat{\psi}}(x)$ and $\widetilde{\tau} = 2\widehat{\tau}_{\mathcal{O}}$, then $|\widetilde{\tau} - \tau| \le 2R_{\widehat{\psi}}(x) + 1$ with the same conditional failure probability. The same statement holds unconditionally after averaging over the even sample.*

Corollary 13 (Finite-library parity validity)**.** *For M even-measurable candidate representations, each satisfying the hypotheses of Proposition 12, and any even-data-measurable selector, simultaneous odd-sample bounds hold with x replaced by $x + \log M$ and total failure probability at most $CT_{\mathcal{O}}e^{-x}$; the selected representation inherits its corresponding split-transfer radius.*

Our parity construction gives held-out validity: it permits selecting a finite library on the even subsequence without reusing the odd data, in the spirit of order-preserving sample splitting (Zou et al., 2020) and free of the prediction-loss pitfalls documented by Pein & Shah (2025).

6 EMPIRICAL EVALUATION

We organize the study as stress tests keyed to the theorem parameters. Unless stated otherwise, $n = 96$, $T = 160$, $\tau = 64$, baseline probability 0.2, and 20% scan trim. We use 200 replicates for primary synthetic cells, 100 for the canonical package audit, and 300 for the null diagnostic. A uniform split over the scan has MAE 26.89 and the fixed midpoint MAE is 16. We record every data-generating mechanism in Appendix G.1.

6.1 SIGNAL STRENGTH AND RETAINED-SIGNAL TRANSITIONS

In Figure 1 we sweep $\delta^2 \in \{0.1, 0.2, 0.4, 0.6, 1, 1.8, 3\}$ over block-aligned, Chung–Lu, and degree-blind changes. At $\delta^2 = 0.6$, Edge U has MAE about 24–26, while the aligned summaries are below 5; at $\delta^2 = 3$, Edge U recovers to about 5–6 and aligned summaries are below 0.25. Degree U remains near the null regime for the degree-blind construction because its mean change is exactly zero. We see that aligned compression moves the localization curve into a substantially easier regime, while a representation that erases the jump gains nothing from its small dimension; Edge U and Wang–Shao nearly coincide, as their algebraic relationship predicts (exact values in Appendix G.2).

Our exact-retention sweeps then fix total jump energy and vary $a_F = \|F\Delta\|^2/\|\Delta\|^2$. With unit jumps $u_\parallel$ in the row space and $u_\perp$ in the null space of the block map, $\Delta(a) = \sqrt{\delta^2/2}\{\sqrt{a}\, u_\parallel + \sqrt{1-a}\, u_\perp\}$

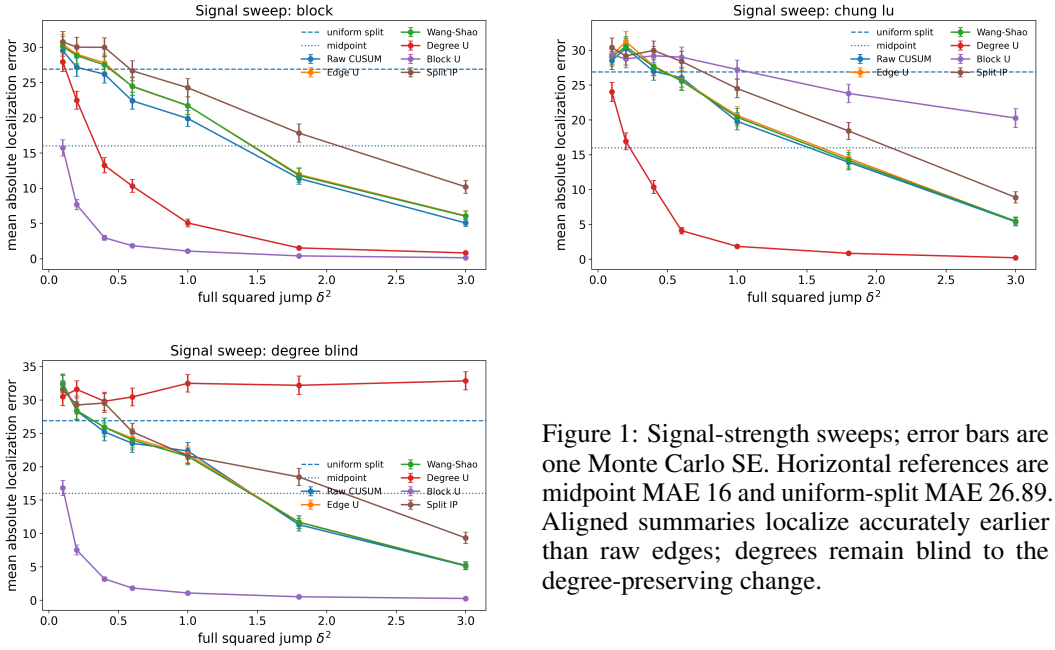


Figure 1: Signal-strength sweeps; error bars are one Monte Carlo SE. Horizontal references are midpoint MAE 16 and uniform-split MAE 26.89. Aligned summaries localize accurately earlier than raw edges; degrees remain blind to the degree-preserving change.

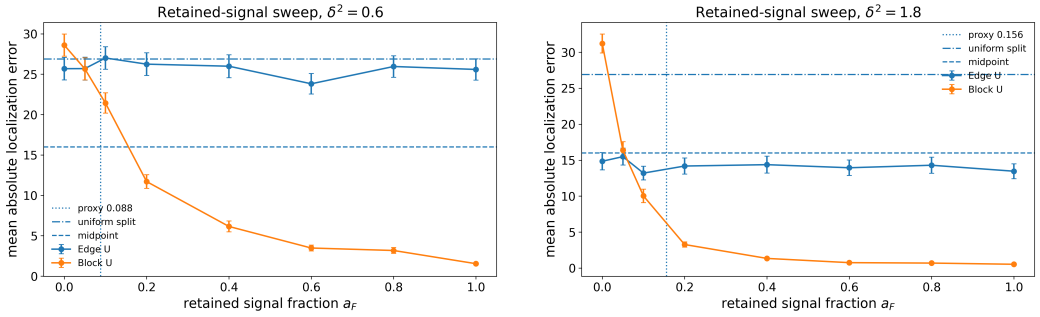


Figure 2: Exact retained-fraction sweeps at moderate signal. The structured map first loses from signal deletion and then wins once retention offsets the reduction in stochastic complexity.

gives exactly $a_F = a$ (Figures 2 and 3). At moderate signal ($\delta^2 = 0.6, 1.8$), we find that Block U overtakes Edge U around $a_F = 0.05$–0.10. At strong signal ($\delta^2 = 3, 5, 8$), low-retention compression becomes clearly harmful and the empirical crossover shifts rightward, matching the direction of the sufficient proxy in Appendix F; we report paired differences and proxy markers in Appendix G.

6.2 Estimated structure and honest finite-library selection

We test Corollary 5 and Lemma 9 with an independent pilot. The pilot shares a four-block partition across regimes, straddles the change, stays independent of the analysis sequence, and fixes the block map by spectral clustering on $T_0 \in \{5, 10, 20, 40\}$ snapshots (Figure 4). We find that at $\delta^2 = 0.6$, raising pilot size T_0 from 5 to 40 raises mean retained fraction from 0.042 to 0.789 and lowers Pilot Block U MAE from 24.50 to 1.95 (oracle 1.34); the analogous $\delta^2 = 1.8$ MAE falls from 13.55 to 0.60 (oracle 0.36). At $T_0 = 20$, using $K = 3, 4, 5$ rather than the true $K = 4$ remains stable: MAEs are $(7.53, 9.34, 8.34)$ for $\delta^2 = 0.6$ and $(1.36, 1.18, 1.21)$ for $\delta^2 = 1.8$. Controlled label corruption verifies the direction of Lemma 9.

We next exercise the parity construction (Proposition 12, Corollary 13). Our candidate library contains edges, degrees, and spectral block maps with $K \in \{2, 3, 4, 5, 6\}$. For a selection sample of length T_s, candidate c has maximizing split $\widehat{k}_c$, maximum corrected-U value $U_c^{\max}$, dimension d_c, and

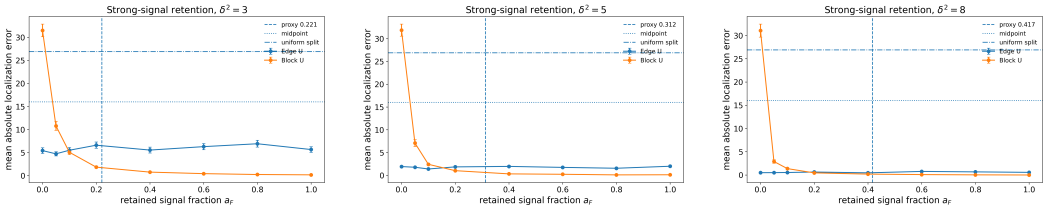


Figure 3: Strong-signal exact-retention stress test at $\delta^2 = 3, 5, 8$ (200 paired replicates per point). Compression is harmful when too little signal survives and becomes useful as a_F increases. Vertical markers are unit-constant sufficient-proxy crossings, not fitted phase boundaries.

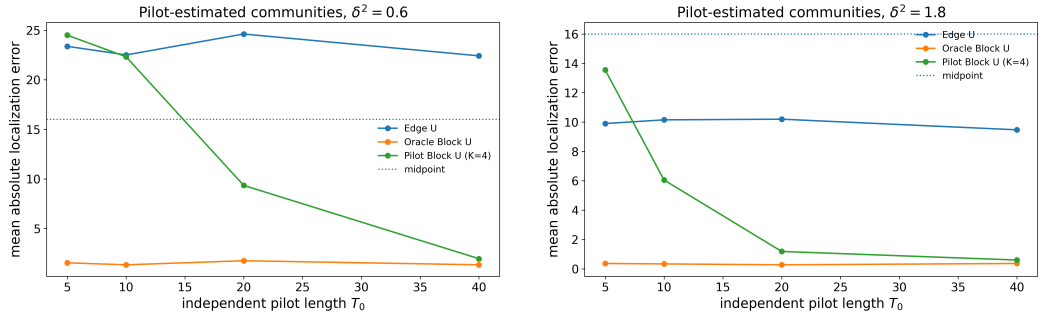


Figure 4: Learning the block partition from an independent pilot at $\delta^2 = 0.6$ (left) and 1.8 (right). Better pilot structure retains more signal and approaches the oracle localization error.

$x_s = \log(eT_s)$. The prespecified score is

$$\widehat{\kappa}^2_{c,\mathrm{raw}} = \max\left\{\frac{T_s U_c^{\max}}{\widehat{k}_c(T_s - \widehat{k}_c)}, 0\right\}, \qquad f_c = 0.30\sqrt{\frac{d_c + x_s}{T_s}}, \tag{12}$$

$$\widehat{\kappa}^2_c = \max\{\widehat{\kappa}^2_{c,\mathrm{raw}} - f_c, 10^{-8}\}, \qquad \widehat{R}_c = \frac{0.30\, x_s}{\widehat{\kappa}^2_c} + \frac{0.30^2 x_s(d_c + x_s)}{T_s(\widehat{\kappa}^2_c)^2}, \qquad \widehat{c} = \arg\min_c \widehat{R}_c. \tag{13}$$

We fixed the single constant 0.30 globally before the simulation loop and use it for every design and signal level; E5B multiplies only f_c by $0.5, 1, 2$. The parity arm ($T_s = 80$) selects by (12) on the even snapshots, freezes the chosen map, and localizes on the odd snapshots before doubling the estimate; the independent-pilot arm selects on a separate length-80 sequence and localizes on all $T = 160$ snapshots; the same-sample arm uses $T_s = 160$ without splitting. The oracle uses degrees for Chung–Lu and the true four-block map otherwise. We report MAE, oracle-family selection frequency, and paired parity-minus-pilot split cost in Table 2. We see that block and degree-blind selection matches the oracle family in at least 99.5% of pilot/parity replicates, and parity localization then improves sharply on raw edges even at $\delta^2 = 0.6$. Chung–Lu exposes the weak-signal limitation: parity chooses degrees only 2.5%, 46.5%, and 84.0% of the time as δ^2 rises. The independent pilot is generally more accurate because selection and localization use separate samples without halving the localization sequence. The E5B floor sweep is stable for block/degree-blind at moderate and strong signal but materially affects Chung–Lu when the dimension floor is aggressive (Appendix G). We report this diagnostic sweep in Table 6 and keep the globally fixed default selector.

6.3 Canonical raw-input baselines and null diagnostics

Our package audit on raw edge vectors uses `gSeg`, `kerSeg`, network WBS, and LAD with 100 replicates per cell (full configuration in Appendix G). The earlier RBF and MST scans ran on a 64-bucket CountSketch, so they remain representation ablations (Appendix G.5) beside this raw-input audit (Table 3). We find that `kerSeg` recovers strongly by $\delta^2 = 3$ (MAE 5.78, 7.17, 4.48 across

Table 2: Finite-library selection over 200 paired replicates: localization MAE (MCSE), oracle-family match (%), and paired parity−pilot error (MCSE).

		Localization MAE					Oracle-family match (%)			Split cost
Design	δ^2	Edge	Oracle	Pilot	Parity	Same	Pilot	Parity	Same	Parity−pilot
Block	.6	24.48 (1.31)	2.13 (.26)	2.19 (.26)	4.75 (.50)	2.21 (.26)	100.0	100.0	100.0	2.56 (.45)
Block	1.8	12.77 (1.09)	.45 (.06)	.44 (.06)	.99 (.15)	.44 (.06)	100.0	100.0	100.0	.55 (.14)
Block	3	6.55 (.69)	.14 (.03)	.19 (.03)	.54 (.08)	.17 (.03)	100.0	100.0	100.0	.35 (.08)
Chung–Lu	.6	25.59 (1.37)	4.91 (.57)	26.45 (1.36)	27.12 (1.31)	23.95 (1.30)	1.0	2.5	4.5	.67 (1.78)
Chung–Lu	1.8	12.51 (.98)	.58 (.09)	6.43 (.83)	8.94 (.97)	1.02 (.20)	39.0	46.5	85.0	2.51 (1.31)
Chung–Lu	3	6.17 (.60)	.24 (.05)	.53 (.11)	2.03 (.41)	.30 (.06)	85.0	84.0	94.0	1.50 (.43)
Degree-blind	.6	23.00 (1.24)	2.15 (.25)	1.98 (.22)	4.58 (.45)	2.11 (.26)	99.5	99.5	100.0	2.60 (.47)
Degree-blind	1.8	12.25 (1.04)	.32 (.05)	.29 (.05)	.54 (.08)	.30 (.05)	100.0	100.0	100.0	.24 (.07)
Degree-blind	3	7.08 (.77)	.10 (.03)	.10 (.03)	.29 (.06)	.09 (.02)	100.0	100.0	100.0	.18 (.05)

Table 3: Canonical raw-input package audit: MAE (MCSE), 100 replicates. Local refinement is excluded from the headline comparison after the E6B threshold-sensitivity diagnostic.

Design	δ^2	gSeg MST	kerSeg	Network WBS	LAD
Block	.6	28.67 (2.08)	26.56 (1.95)	32.08 (2.12)	27.50 (1.86)
Block	1.8	29.16 (2.09)	12.08 (1.40)	21.06 (1.96)	26.13 (1.72)
Block	3	26.65 (2.05)	5.78 (.87)	21.00 (2.12)	25.82 (1.83)
Chung–Lu	.6	30.66 (1.92)	27.18 (1.97)	31.12 (2.47)	30.53 (1.74)
Chung–Lu	1.8	30.11 (1.94)	16.78 (1.81)	28.40 (2.49)	24.70 (1.80)
Chung–Lu	3	33.25 (1.99)	7.17 (1.15)	17.50 (1.92)	28.76 (1.79)
Degree-blind	.6	29.85 (1.94)	27.30 (1.97)	32.64 (2.27)	27.93 (1.75)
Degree-blind	1.8	30.22 (2.04)	14.52 (1.59)	21.18 (1.94)	26.14 (1.67)
Degree-blind	3	29.46 (2.04)	4.48 (.79)	19.06 (2.20)	27.01 (1.81)

Table 4: Prespecified strong-signal package sanity check (50 fresh replicates per cell): MAE (MCSE). Refinement is initialized at the true change. The diagnostic validates WBS but shows substantial threshold sensitivity of local refinement.

Design	δ^2	WBS	Refine default	Refine $0.5\tau_2$	gSeg 5-NN
Block	8	1.00 (.36)	16.68 (2.01)	5.20 (.98)	6.24 (1.43)
Block	20	.00 (.00)	.08 (.06)	.40 (.16)	7.04 (1.76)
Chung–Lu	8	2.04 (.53)	30.00 (.00)	6.32 (1.35)	9.08 (1.91)
Chung–Lu	20	.04 (.04)	19.32 (1.96)	.44 (.17)	1.10 (.23)
Degree-blind	8	1.60 (.47)	26.92 (1.20)	6.44 (1.10)	5.20 (1.00)
Degree-blind	20	.04 (.04)	.16 (.08)	.40 (.18)	1.00 (.23)

the three designs). Our prespecified E6B sanity check validates WBS at stronger signals: strongest-candidate and thresholded WBS agree, with MAE $(1.00, 2.04, 1.60)$ at $\delta^2 = 8$ and $(0.00, 0.04, 0.04)$ at $\delta^2 = 20$. Table 4 summarizes this check, and Table 7 adds the thresholded-WBS and single-MST columns. Local refinement remains highly threshold-sensitive even from the true initializer, so we report it as a diagnostic; a denser 5-NN graph also substantially improves `gSeg` over the single MST. We see that the anomaly is a threshold effect; rather than select a favorable threshold after observing it, we drop refinement from the headline comparison.

Our null diagnostics compare deliberately uninformative representations under change and no change. We find that Degree U under the degree-blind change has MAE 32.19 and outer-quartile mass 0.663, close to its no-change control $(30.49,\ 0.610)$; the frozen six-eigenvalue Spectral U on a block change behaves similarly (30.62/0.630 vs. 31.65/0.613 under no change). We see that the below-uniform MAE in these cells reflects a boundary-concentrated, null-like argmax distribution. We give full histograms in Appendix G; median absolute errors agree (Table 8).

6.4 Real data: Calendar robustness on Enron email-data

We project the public temporal Enron email data of Benson et al. (2018) to undirected pairs among 143 core addresses ($D = 10{,}153$). The original July–December analysis gives Sep. 22 for edges

Table 5: Enron calendar robustness in the July–November window. Ranges are descriptive fitted-regime resampling percentiles, not confidence intervals.

Calendar	Edges: estimate (descriptive range)	Degrees: estimate (descriptive range)
Daily	Sep. 25 (Sep. 19–26)	Sep. 25 (Sep. 17–26)
Weekdays	Sep. 25 (Sep. 19–25)	Sep. 25 (Sep. 17–26)
Weekly	week Sep. 17 (Sep. 10–17)	week Sep. 17 (Sep. 3–17)

and Sep. 25 for degrees. Because Sep. 22, 2001 was a Saturday and the six-month window also contains later corporate events, we added a shorter window once that concern arose and co-report both analyses with the same two predetermined representations. Our shorter July 1–Nov. 15 robustness window targets the weekend concern without using historical event dates in tuning. We find that every daily and weekday-only analysis across 36 window/trim combinations selects Sep. 25 for both representations. Weekly sensitivity lies in the weeks beginning Sep. 17 or Sep. 24 (median Sep. 24), consistent with coarser bins that can straddle a daily transition. We reproduce the original July–December result exactly. Whole-snapshot permutations use 999 random orders, and we treat fitted-regime ranges as descriptive. We see that the 71-fold degree compression preserves the strongest late-September feature under weekend-targeted calendar controls (Table 5). We give full construction and resampling details in Appendix G.

7 Discussion

Our central claim is two-sided. Over unrestricted Bernoulli changes, unknown direction creates an intrinsic $D\rho^2/(T\delta^4)$ localization cost; valid structural side information can reduce that term when it rules out directions without sacrificing change signal. Our upper bounds make the same tradeoff operational through (d_F, κ_F): known groups are useful when scientifically meaningful interactions are expected to move coherently across those groups; degrees are attractive when node-level activity is the relevant changing object; and raw edges remain the natural safeguard when no such alignment is credible. Our results also explain why a highly compressed representation can perform poorly while structure remains valuable: compression is useful only when it removes more stochastic complexity than change information. The retained-signal fraction a_F and the sufficient upper-bound comparison in Appendix F make this advice quantitative.

Our empirical results mirror the same hierarchy. Signal sweeps show that raw methods recover once the jump is strong, while aligned structure shifts the localization curve into substantially weaker signal regimes. Exact-a_F sweeps display both sides of the tradeoff: at strong signal, low retention is materially harmful and the crossover point moves rightward. Pilot-learned communities approach oracle performance as misclustering decreases and remain stable to one-block under/over-specification. Parity selection provides honest finite-library validity but pays a split cost and can over-penalize the higher-dimensional degree family at weak Chung–Lu signal; our E5B sensitivity analysis makes that limitation explicit. The raw-input package audit removes the earlier CountSketch confound: `kerSeg` recovers at strong signal, E6B validates WBS, and we keep local refinement out of the headline table because its outcome is threshold-sensitive. Null diagnostics separate representation blindness from a stable wrong target, while Enron provides a descriptive calendar-robust illustration.

Limitations. The structural lower bound is aligned-aggregation specific; a general non-aligned lower bound is open. The s-sparse analogue of Theorem 6 is not developed. The misclustering lemma is worst-case and can be loose when misclustered vertices carry little jump mass. Proposition 10 has no matching n-term lower bound over the whole Chung–Lu class. The nonlinear theorem is a bounded frozen-feature transfer result without a positive learned-encoder example. The parity construction is not same-data oracle adaptation. Multiple changes, temporal dependence, and formal real-data uncertainty remain for future work. Appendix F.5 collects the full scope conditions. Structure buys better localization only through change signal that the representation actually preserves.

Reproducibility Statement

The supplementary repository contains the codes for fixed-representation estimators, notebooks/scripts for all experiments and diagnostics, frozen per-replicate outputs, seeds, and manifests, package-audit code, and figure/table regeneration files. Appendices A–E provide complete proofs and state the assumptions used by the theoretical claims, while Appendix G records the synthetic mechanisms, independent-pilot and parity-selection constructions, canonical package settings, null diagnostics, and the Enron calendar/resampling protocol. Raw Enron data are not redistributed; the repository identifies the public Benson/Cornell source and reconstructs the reported analysis from it.

AI Use Statement

Generative AI tools assisted with manuscript restructuring and drafting, proof organization/checking, experiment-code development/debugging, archived-output auditing, literature-oriented drafting, and reproducibility packaging. Reported simulations and real-data runs were executed by the authors in the recorded environments; the authors reviewed the text, mathematical claims, citations, code, and numerical summaries and take responsibility for the final work.

References

Jushan Bai. Common breaks in means and variances for panel data. *Journal of Econometrics*, 157(1): 78–92, 2010. doi: 10.1016/j.jeconom.2009.10.020.

Austin R. Benson, Rediet Abebe, Michael T. Schaub, Ali Jadbabaie, and Jon Kleinberg. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences*, 115(48):E11221–E11230, 2018. doi: 10.1073/pnas.1800683115. URL https://www.cs.cornell.edu/~arb/data/email-Enron/.

Monika Bhattacharjee, Moulinath Banerjee, and George Michailidis. Change point estimation in a dynamic stochastic block model. *Journal of Machine Learning Research*, 21(107):1–59, 2020. URL https://jmlr.org/papers/v21/18-814.html.

Hao Chen and Nancy R. Zhang. Graph-based change-point detection. *The Annals of Statistics*, 43(1): 139–176, 2015. doi: 10.1214/14-AOS1269.

Song Xi Chen and Ying-Li Qin. A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835, 2010. doi: 10.1214/09-AOS716. URL https://arxiv.org/abs/1002.4547.

Haeran Cho and Housen Li. Covariance scanning for adaptively optimal change point detection in high-dimensional linear models, 2025. arXiv:2507.02552, v4 (1 June 2026).

Damien Garreau and Sylvain Arlot. Consistent change-point detection with kernels. *Electronic Journal of Statistics*, 12(2):4440–4486, 2018. doi: 10.1214/18-EJS1513.

Evarist Giné, Rafał Latała, and Joel Zinn. Exponential and moment inequalities for U-statistics. In Evarist Giné, David M. Mason, and Jon A. Wellner (eds.), *High Dimensional Probability II*, volume 47 of *Progress in Probability*, pp. 13–38. Birkhäuser, Boston, MA, 2000. doi: 10.1007/978-1-4612-1358-1_2.

Christian Houdré and Patricia Reynaud-Bouret. Exponential inequalities, with constants, for U-statistics of order two. In Evarist Giné, Christian Houdré, and David Nualart (eds.), *Stochastic Inequalities and Applications*, volume 56 of *Progress in Probability*, pp. 55–69. Birkhäuser, Basel, 2003. doi: 10.1007/978-3-0348-8069-5_5.

Shenyang Huang, Yasmeen Hitti, Guillaume Rabusseau, and Reihaneh Rabbany. Laplacian change point detection for dynamic graphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 349–358, 2020. doi: 10.1145/3394486.3403077.

Abhishek Kaul, Stergios B. Fotopoulos, Venkata K. Jandhyala, and Abolfazl Safikhani. Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics*, 15 (1):71–134, 2021. doi: 10.1214/20-EJS1791.

Yik Lun Kei, Jialiang Li, Hangjian Li, Yanzhen Chen, and Oscar Hernan Madrid Padilla. Change point detection in dynamic graphs with decoder-only latent space model. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=DVeFqV56Iz. arXiv:2404.04719.

Haoyang Liu, Chao Gao, and Richard J. Samworth. Minimax rates in sparse, high-dimensional change point detection. *The Annals of Statistics*, 49(2):1081–1112, 2021. doi: 10.1214/20-AOS1994.

Oscar Hernan Madrid Padilla, Yi Yu, and Carey E. Priebe. Change point localization in dependent dynamic nonparametric random dot product graphs. *Journal of Machine Learning Research*, 23 (234):1–59, 2022. URL https://jmlr.org/papers/v23/20-643.html.

Leto Peel and Aaron Clauset. Detecting change points in the large-scale structure of evolving networks. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, volume 29, pp. 2914–2920, 2015. doi: 10.1609/aaai.v29i1.9574.

Florian Pein and Rajen D. Shah. Cross-validation for change-point regression: pitfalls and solutions. *Bernoulli*, 31(1):388–411, 2025. doi: 10.3150/24-BEJ1732.

Emmanuel Pilliat, Alexandra Carpentier, and Nicolas Verzelen. Optimal multiple change-point detection for high-dimensional data. *Electronic Journal of Statistics*, 17(1):1240–1315, 2023. doi: 10.1214/23-EJS2126.

Déborah Sulem, Henry Kenlay, Mihai Cucuringu, and Xiaowen Dong. Graph similarity learning for change-point detection in dynamic networks. *Machine Learning*, 113(1):1–44, 2024. doi: 10.1007/s10994-023-06405-x.

Nicolas Verzelen, Magalie Fromont, Matthieu Lerasle, and Patricia Reynaud-Bouret. Optimal change-point detection and localization. *The Annals of Statistics*, 51(4):1586–1610, 2023. doi: 10.1214/23-AOS2297.

Daren Wang, Yi Yu, and Alessandro Rinaldo. Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics*, 49(1):203–232, 2021. doi: 10.1214/20-AOS1953. URL https://arxiv.org/abs/1809.09602.

Runmin Wang and Xiaofeng Shao. Dating the break in high-dimensional data. *Bernoulli*, 29(4):2879–2901, 2023. doi: 10.3150/22-BEJ1567. URL https://arxiv.org/abs/2002.04115.

Tengyao Wang and Richard J. Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1): 57–83, 2018. doi: 10.1111/rssb.12243. URL https://academic.oup.com/jrsssb/article/80/1/57/7048412.

Yi Yu. A review on minimax rates in change point detection and localisation, 2020. arXiv:2011.01857.

Daniele Zambon, Cesare Alippi, and Lorenzo Livi. Change-point methods on a sequence of graphs. *IEEE Transactions on Signal Processing*, 67(24):6327–6341, 2019. doi: 10.1109/TSP.2019.2953596.

Changliang Zou, Guanghui Wang, and Runze Li. Consistent selection of the number of change-points via sample-splitting. *The Annals of Statistics*, 48(1):413–439, 2020. doi: 10.1214/19-AOS1814.

A Population identities and variance drift

A.1 Proof of Proposition 1

For a block of size m with means $\mu_1, \dots, \mu_m$ and average $\bar{\mu}$, by independence,

$$\mathbb{E}\frac{\sum_{s\neq t}\langle X_s, X_t\rangle}{m(m-1)} = \|\bar{\mu}\|^2 - \frac{\sum_t \|\mu_t - \bar{\mu}\|^2}{m(m-1)}.$$

The cross-block average has expectation equal to the inner product of the two block mean averages. At $k = \tau$ both blocks are pure, so the subtracted quantities vanish. For $k > \tau$ only the left block is mixed. Its mean dispersion is $\tau(k-\tau)\kappa^2/k$. Substituting in (3), we obtain the stated expression; for $k < \tau$, we obtain

$$\mathbb{E}U_k = Q_k^0 - \frac{(\tau-k)(T-\tau)k}{T(T-k)(T-k-1)}\kappa^2.$$

The ordinary population CUSUM losses are

$$Q_\tau^0 - Q_k^0 = \begin{cases} \tau(k-\tau)\kappa^2/k, & k > \tau, \\ (T-\tau)(\tau-k)\kappa^2/(T-k), & k < \tau. \end{cases}$$

The mixed-block correction is nonnegative, so these losses prove population separation on the trimmed range. Expanding $\sum_{s\neq t}\langle X_s, X_t\rangle = \|\sum_s X_s\|^2 - \sum_s \|X_s\|^2$ also proves the sample-variance correction identity after (3). □ Because of the mixed-block term, we call U a corrected contrast rather than an unbiased estimate of Q_k^0 at every split.

Exact prefix-sum implementation. We see that the apparent pair sums in (3) need only linear work in time. Let $S_k = \sum_{t\le k} X_t$ and $V_k = \sum_{t\le k}\|X_t\|^2$. Then the first ordered-pair sum equals $\|S_k\|^2 - V_k$, the right-block analogue uses $S_T - S_k$ and $V_T - V_k$, and the cross term is $\langle S_k, S_T - S_k\rangle$. Once the features X_t are formed, we evaluate a full scan from running prefix summaries in $O(Tq)$ arithmetic for a q-dimensional representation. Separately from the localization theorem, we see a practical consequence of structural compression: degrees or known block aggregates reduce the scan dimension as well as the stochastic dimension. Feature construction itself depends on how the graph stream is stored, and we count it separately from any statistical gain.

A.2 Raw CUSUM expectation and inconsistency

Let $w_{k,t}$ denote the normalized CUSUM weights. Each weight is $\sqrt{(T-k)/(Tk)}$ on the left and $-\sqrt{k/[T(T-k)]}$ on the right. The weights sum to zero, their squared sum is one, and independence gives

$$\mathbb{E}Q_k = Q_k^0 + \sum_t w_{k,t}^2 \operatorname{tr}\operatorname{Var}(X_t).$$

For $k > \tau$, the coefficient of v_1 is $\tau(T-k)/(Tk)$. Subtracting its value $(T-\tau)/T$ at the true split and using the population loss above proves (2). By time reversal,

$$\mathbb{E}Q_k - \mathbb{E}Q_\tau = \frac{\tau-k}{T-k}\{v_1 - v_2 - (T-\tau)\kappa^2\}, \qquad k < \tau.$$

Proposition 14 (Raw maximizer can converge to a scan boundary)**.** *Take even $T \to \infty$, $n = T^2$, $\tau = T/2$, $\eta = 1/5$, and homogeneous independent networks with $p_1 = 1/5$ and $p_2 = 1/5 + 2/(5T)$ on every edge. The raw edge maximizer satisfies $\widehat{\tau}_Q/T \to_{\mathbb{P}} 4/5$, while the right side of Corollary 3, with $x \asymp \log T$, tends to zero for U.*

Proof. Write $h = 2/(5T)$. In edge coordinates, $\kappa^2 = Dh^2$ and $v_2 - v_1 = Dh(3/5 - h)$. We see that $v_2 - v_1 - \tau\kappa^2 = Dh(2/5 - h) > 0$ for large T. The expected raw objective increases strictly to the right of τ; to the left its expectation is below the value at τ. Its maximum is at $\lfloor 4T/5\rfloor$.

For every fixed $\epsilon > 0$, the expected loss at any split at least ϵT away from this boundary is at least $c_\epsilon D/T$. For one edge, its centered normalized CUSUM is uniformly sub-Gaussian; its mean is

$O(T^{-1/2})$. Its squared value has bounded variance. Edge independence gives $\mathrm{Var}(Q_k) \le CD$ uniformly. Chebyshev and a union bound show

$$\mathbb{P}\left(\max_k |Q_k - \mathbb{E}Q_k| > \epsilon D/T\right) \le C_\epsilon T^3/D = O(T^{-1}) \to 0.$$

We obtain the boundary convergence. Here $\kappa^2 \asymp T^2$, $D \asymp T^4$ and $\rho \asymp 1$. The corrected bound is $O(\log T/T^2 + \log T/T)$, so it eventually implies exact localization on its high-probability event. □

B VARIANCE-SENSITIVE CONCENTRATION AND FIXED-MAP LOCALIZATION

We first record the precise concentration input. For a matrix M, write $\|M\|_{\max} = \max_{i,j} |M_{ij}|$ and $r_2(M) = \max_i (\sum_j M_{ij}^2)^{1/2}$.

Lemma 15 (Bounded-coordinate quadratic form)**.** *Let $\epsilon_1, \dots, \epsilon_m$ be independent, centered, $|\epsilon_i| \le 1$, and $\mathbb{E}\epsilon_i^2 \le \rho$. If M is symmetric with zero diagonal, then for $x \ge 1$,*

$$\mathbb{P}\{|\epsilon^\top M \epsilon| > C[\rho \|M\|_F \sqrt{x} + \rho \|M\|_{\mathrm{op}} x + \sqrt{\rho}\, r_2(M) x^{3/2} + \|M\|_{\max} x^2]\} \le Ce^{-x}. \quad (14)$$

Proof. We specialize the canonical U-statistic moment inequality in Giné et al. (2000, Theorem 3.3) and give the reduction to identify all four parameters and justify its use for nonidentical Bernoulli variables. Introduce independent selectors $\zeta_i \sim \mathrm{Bern}(1/2)$, independent of ϵ. For every fixed ϵ,

$$\epsilon^\top M \epsilon = 4\mathbb{E}_\zeta \sum_{i:\zeta_i=1} \sum_{j:\zeta_j=0} M_{ij}\epsilon_i\epsilon_j.$$

By Jensen, we bound the L^p norm, $p \ge 2$, by four times the largest L^p norm of such a cross sum. For a fixed partition the two groups of variables are independent. Its kernels $h_{ij}(x,y) = M_{ij}xy$ are canonical because each variable is centered. Zero kernels pad unequal group sizes.

Set $\sigma_i^2 = \mathbb{E}\epsilon_i^2$. The four quantities in the cited theorem are bounded, respectively, by

$$\begin{aligned} C_0^2 &= \sum_{i,j} M_{ij}^2 \sigma_i^2 \sigma_j^2 \le \rho^2 \|M\|_F^2, \\ D_0 &= \|\mathrm{diag}(\sigma) M\, \mathrm{diag}(\sigma)\|_{\mathrm{op}} \le \rho \|M\|_{\mathrm{op}}, \\ B_0^2 &\le \rho \max_i \sum_j M_{ij}^2, \qquad A_0 \le \|M\|_{\max}. \end{aligned}$$

For D_0, Cauchy–Schwarz gives $|\mathbb{E}\epsilon_i f_i(\epsilon_i)| \le \sigma_i (\mathbb{E}f_i^2)^{1/2}$, which identifies the operator acting between the two direct sums of L^2 spaces. All displayed norms dominate the corresponding rectangular submatrix norms. For B_0, use $|x| \le 1$ in $\sum_j \mathbb{E}h_{ij}(x,\epsilon_j)^2$; the column version is identical. We conclude that the pth norm is at most $C(\sqrt{p}C_0 + pD_0 + p^{3/2}B_0 + p^2 A_0)$. Taking $p = \max(2, x)$ and applying Markov's inequality at a constant times this norm proves (14). □

Lemma 16 (Temporal kernel bounds)**.** *Let A_k be symmetric, have diagonal zero, and off-diagonal entries*

$$(A_k)_{st} = \begin{cases} (T-k)/[T(k-1)], & s,t \le k, \\ k/[T(T-k-1)], & s,t > k, \\ -1/T, & \text{otherwise}. \end{cases}$$

Then $A_k \mathbf{1} = 0$ and $U_k = \sum_{s,t} (A_k)_{st} \langle X_s, X_t\rangle$. For $k, \tau \in \mathcal{K}_\eta$, $u = |k-\tau| \ge 1$, $B = A_k - A_\tau$ and $z_t = \mathbf{1}\{t > \tau\}$,

$$\begin{aligned} \|B\|_F, \|B\|_{\mathrm{op}} \le C_\eta \sqrt{u/T}, \qquad r_2(B) \le C_\eta T^{-1/2}, \qquad \|B\|_{\max} \le C_\eta/T, \\ \|Bz\| \le C_\eta \sqrt{u}, \qquad \|Bz\|_\infty \le C_\eta. \end{aligned} \quad (15)$$

Proof. There are u indices changing sides. Entries involving one of these indices change by at most C_η/T, and there are at most $2uT$ such entries. All remaining entries change by at most $C_\eta u/T^2$: apply the mean value theorem to the two within-block coefficients, whose derivatives are bounded by

C_η/T^2 on the trimmed range. Squaring and summing yields $\|B\|_F^2 \le C_\eta u/T$; the operator bound follows. Each row has at most T entries bounded by C_η/T, proving the row and entry bounds. In Bz, a changed row has magnitude at most C_η. An unchanged row has at most u entries of size C_η/T and T entries of size $C_\eta u/T^2$, so has magnitude at most $C_\eta u/T$. Summing squared row bounds gives $\|Bz\|^2 \le C_\eta(u + u^2/T)$. The row-sum identity follows by counting the within-block and cross-block entries in any row. □

Proof of Theorem 2 with the exact range term. Set $m_\Delta = \|H\Delta\|_\infty = m_F(\Delta)$. Write $Y_t = p_1 + z_t\Delta + \epsilon_t$, stacking errors over time and edges. Since $B\mathbf{1} = 0$, the common level cancels, and expanding we obtain

$$(U_k - U_\tau) - \mathbb{E}(U_k - U_\tau) = 2\sum_t (Bz)_t\langle H\Delta, \epsilon_t\rangle + \epsilon^\top(B \otimes H)\epsilon. \tag{16}$$

The quadratic form has zero expectation because B has zero diagonal and different edge-time coordinates are independent. Since $0 \preceq H \preceq I$,

$$\|H\Delta\| \le \kappa, \quad \|H\|_F \le \sqrt{d}, \quad \|H\|_{\mathrm{op}} \le 1, \quad r_2(H) \le \sqrt{\ell}, \quad \|H\|_{\max} \le \ell.$$

The last two inequalities follow from $H^2 \preceq H$ and $|H_{ef}| \le \sqrt{H_{ee}H_{ff}}$. Bernstein's inequality, using Lemma 16, bounds the linear term by $C_\eta\{\kappa\sqrt{\rho u x} + m_\Delta x\}$. Lemma 15 applied to $B \otimes H$ bounds the quadratic term by

$$C_\eta\left\{\rho(\sqrt{dx} + x)\sqrt{u/T} + \sqrt{\rho\ell/T}\, x^{3/2} + \ell x^2/T\right\}.$$

A union bound over at most T splits and the population loss $c_\eta u\kappa^2$ from Proposition 1 give (5), which is exactly the stated bound since $m_\Delta = m_F(\Delta)$. Using (4) and $T\rho \ge \ell x$, we then obtain the leverage simplification in (6). □

Proof of Corollary 3. Use $m_F(\Delta) \le C_0\rho$ in Theorem 2; the same aggregation condition absorbs the $x^{3/2}$ and x^2 terms. □

Proof of Corollary 4. Take $x = c_0\log(eT)$ in Corollary 3 with c_0 large enough that $CTe^{-x} \le T^{-2}$. On the complement of the high-probability event use the trivial bound $|\widehat{\tau}_U - \tau| \le T$. □

When the leverage simplification matters. Inequality (4) follows from $m_F(\Delta) = \max_e |\langle Fe_e, F\Delta\rangle|$, Cauchy–Schwarz, and $H_{ee} = \|Fe_e\|^2$. The leverage replacement can be slack for dense projections because it maximizes over jump directions. Moreover, if $\rho \ge \ell x$, then the worst-case term $\sqrt{\ell}x/\kappa$ is never larger than both the scalar term $\rho x/\kappa^2$ and one time unit: when $\kappa \le \rho/\sqrt{\ell}$ the scalar term dominates, while otherwise $\sqrt{\ell}x/\kappa < \ell x/\rho \le 1$. We see that above unit localization resolution the leverage correction can matter only in the sparser regime $\rho < \ell x$; Appendix F.5 records the open question of its necessity there. Theorem 2 permits dense orthogonal embeddings; when their leverage is small, the additional term is itself small.

B.1 Constant-mean testing

Under a constant mean, the linear term is absent and A_k itself has Frobenius and operator norms bounded by C_η, row norm at most $C_\eta/\sqrt{T}$, and maximum entry at most C_η/T. We see that when $T\rho \ge \ell x$,

$$\mathbb{P}\{|U_k| > C_\eta\rho(\sqrt{dx} + x)\} \le Ce^{-x}.$$

For a predetermined finite set of interval–split pairs, the analogous bound uses each interval's length and trimming. A union bound controls the family-wise error if $x \ge \log(CN_{\mathrm{scan}}/\beta)$ and the aggregation condition holds on every interval. We thereby obtain family-wise null control for the scan.

Under the independent constant-mean network model, snapshots are exchangeable. We also give the finite randomization p-value

$$\frac{1 + \#\{b : S(Y^{\pi_b}) \ge S(Y)\}}{B + 1}, \qquad S(Y) = \max_{k \in \mathcal{K}_\eta} U_k,$$

with independent uniform time permutations π_b. It is super-uniform: conditionally on a permutation orbit, the observed sequence is uniform on that orbit, and the B independently permuted sequences are independent uniform orbit elements. The rank is uniform without ties and the weak inequality is conservative with ties.

C Proof of the two-term lower bound

C.1 Proof sketch

We use baseline $p = 3\rho/4$ and post-change probabilities $p + aS_e$, with independent signs $S_e \in \{-1, 1\}$ and $a = \delta/\sqrt{2D}$. Set $c = a^2/\{p(1-p)\}$. For two locations u apart, condition on their common post-change suffix of length b. If m is the posterior mean of one sign, then $\mathbb{E}m^2 \le 2bc$. The disputed block has conditional chi-square divergence

$$\frac{(1+c)^u + (1-c)^u}{2} - 1 + \frac{m^2}{2}\{(1+c)^u - (1-c)^u\}.$$

When $Tc \le 1$, summing over edges we obtain $\mathrm{KL}(M_0 \| M_1) \le CDuTc^2$, and two-point testing with $u \asymp \min\{T, 1/(DTc^2)\}$ gives the second term. For $Tc > 1$, a fixed direction supplies the first term instead. Random-sign mixtures are standard in high-dimensional detection lower bounds (Liu et al., 2021; Pilliat et al., 2023); our localization-specific step is to compare two nearby change locations while conditioning on their common suffix, so that posterior uncertainty about the random signs survives into the disputed time block and yields the second localization term away from the detection boundary. We impose the lower cutoff $R_D \ge C_\eta$ in Theorem 6 so that the polynomial lower bound stays above the integer resolution of the experiment, and Corollary 4 puts the upper bound on the same expected-risk scale. When $D \lesssim \log(eT)$, the $\log^2(eT)$ stochastic floor in the high-probability upper bound dominates its D-term.

C.2 Full proof

Set $p = 3\rho/4$, $a = \delta/\sqrt{2D}$, $v = p(1-p)$ and $c = a^2/v$. Then $a \le \rho/4$, $v \asymp \rho$, and $p, p \pm a$ lie in $[\rho/2, \rho]$. Define

$$A = \frac{1}{Dc}, \qquad B = \frac{1}{DTc^2}.$$

These are comparable to the scalar and dimension-dependent terms in (9). Suppose first $Tc \le 1$. For each edge independently draw $S_e \in \{-1, 1\}$ uniformly, held fixed across the sequence. Conditional on the signs, the pre-change probability is p and the post-change probability is $p + aS_e$. Every member of this mixture has the same full jump δ.

We compare locations $\tau_0 < \tau_1$ around $T/2$, separated by u. Both locations will remain in $\mathcal{K}_\eta$ by choosing $u \le c_\eta T$. For one edge, the common suffix $R = (Y_{\tau_1+1}, \dots, Y_T)$, of length $b = T - \tau_1$, has the same mixture law under both hypotheses. Let $m(R) = \mathbb{E}(S_e \mid R)$. Under the later-change law the disputed block of length u is i.i.d. $\mathrm{Bern}(p)$ and independent of R. Under the earlier-change law, its conditional density relative to that baseline is

$$L_m = \tfrac{1}{2}(1+m)L_+ + \tfrac{1}{2}(1-m)L_-,$$

where $L_\pm$ are u-sample likelihood ratios for $p \pm a$ against p. For one observation, direct summation shows $\mathbb{E}_p L_s L_{s'} = 1 + ss'c$, and taking products and expanding the mixture we obtain

$$\chi^2(L_m, p^{\otimes u}) = \tfrac{1}{2}\{(1+c)^u + (1-c)^u\} - 1 + \tfrac{1}{2}m^2\{(1+c)^u - (1-c)^u\}. \tag{17}$$

From the even and odd binomial terms, for $uc \le 1$,

$$\tfrac{1}{2}\{(1+c)^u + (1-c)^u\} - 1 \le \tfrac{1}{2}(uc)^2 e^{uc}, \quad \tfrac{1}{2}\{(1+c)^u - (1-c)^u\} \le uc\, e^{uc}.$$

Pinsker applied to the posterior sign law versus its uniform prior gives $\mathbb{E}m(R)^2 \le 2I(S_e; R)$. For any reference law Q_R,

$$I(S_e; R) \le \mathbb{E}_{S_e} \mathrm{KL}(\mathcal{L}(R \mid S_e) \| Q_R).$$

Take $Q_R = \mathrm{Bern}(p)^{\otimes b}$. The Bernoulli bound $\mathrm{KL}(\mathrm{Bern}(p \pm a) \| \mathrm{Bern}(p)) \le a^2/v = c$ gives $\mathbb{E}m(R)^2 \le 2bc$.

The common prefix is independent of the sign and cancels from the likelihood ratio. Conditioning on R, using $\mathrm{KL} \le \chi^2$, and averaging (17), we bound the one-edge divergence by

$$e^{uc}c^2(u^2/2 + 2bu) \le CuTc^2.$$

The edge mixtures are independent, so $\mathrm{KL}(M_0 \| M_1) \le CDuTc^2$. We choose $u = \lfloor c_0 \min(T, B) \rfloor$; a small constant c_0 makes the divergence at most $1/8$ and respects trimming. Since $Tc \le 1$, $B \ge A$, so $u \asymp \min(T, A + B)$. The assumption $R_D \ge C_\eta$, with a large enough constant, ensures that this comparison survives integer rounding.

If $Tc > 1$, we fix every sign to $+1$. The laws at the two locations differ only on u snapshots, and their divergence is at most Duc. Choose $u = \lfloor c_0 \min(T, A) \rfloor$ to make it at most $1/8$. Now $B < A$, so again $u \asymp \min(T, A + B)$.

In both regimes, $\mathrm{TV} \le 1/4$. Turning any location estimate into a test by choosing the nearer of the two locations, we see that its average absolute error under the two laws is at least $u(1 - \mathrm{TV})/4 \ge 3u/16$. A mixture risk is an average of risks of admissible fixed parameters. The supremum in (9) dominates this average, proving Theorem 6. □

D Structural proofs

D.1 Feature constants

The four maps of Table 1 are $FY = Y$ (edges), $\mathbf{1}^\top Y/\sqrt{D}$ (count), $GY/\sqrt{2(n-1)}$ (degrees), and $(\sum_{e \in b} Y_e/\sqrt{N_b})_b$ (blocks), where N_b counts edges in a known block pair. For edge coordinates, $F = I_D$ gives $H = I_D$ and $\ell = 1$. For normalized count, $F = \mathbf{1}^\top/\sqrt{D}$ gives $H = \mathbf{1}\mathbf{1}^\top/D$, $d = 1$, and $\ell = 1/D$. For the unsigned vertex–edge incidence matrix G, every edge column contains exactly two ones, so $GG^\top = (n-2)I + \mathbf{1}\mathbf{1}^\top$, whose largest eigenvalue is $2(n-1)$ and whose smallest is $n - 2$. For $n \ge 3$, the degree map has rank n and operator norm one. In $H = G^\top G/[2(n-1)]$, the diagonal is $1/(n-1)$. Entries are nonnegative and every row sums to one, so $m_F(\Delta) \le \rho$. The feature mean difference is precisely $(P_2 - P_1)\mathbf{1}/\sqrt{2(n-1)}$.

For a block pair with N_b edges the feature row has entries $N_b^{-1/2}$ on that pair and zero elsewhere. Disjoint supports give orthonormal rows. H has entries $1/N_b$ within pair b and zero between pairs; its row sums are one and $\ell = 1/N_{\min}$. If a jump is constant with value a_b on the pair, both its edge squared norm and its feature squared norm equal $\sum_b N_b a_b^2$.

D.2 Proof of Theorem 7 and Corollary 8

For cell b, let the feature row have entries $N_b^{-1/2}$ on $\mathcal{E}_b$ and zero elsewhere. The rows have disjoint supports and are orthonormal. A cell-constant jump is preserved in norm, so $\kappa^2 = \delta^2/2$. We obtain the upper bound from Corollary 3, with d cells and $\ell = 1/N_{\min}$. For the lower bound, we use baseline $p = 3\rho/4$ and independently draw one sign S_b per cell. The post-change probability on cell b is $p + a_b S_b$, where

$$a_b = \frac{\delta}{\sqrt{2dN_b}} \le \rho/4.$$

Every parameter has full jump δ. Set $c_b = a_b^2/[p(1-p)]$ and $\widetilde{c} = N_b c_b = \delta^2/[2dp(1-p)]$. The disputed time block supplies uN_b Bernoulli observations sharing the sign and the common suffix supplies $b_{\mathrm{time}} N_b$ observations. In (17) replace u by uN_b and c by c_b. When $T\widetilde{c} \le 1$, the same conditional-mixture calculation gives one-cell divergence at most $CuT\widetilde{c}^2$ and total divergence at most $CduT\widetilde{c}^2$. When $T\widetilde{c} > 1$, fixed signs give divergence at most $du\widetilde{c}$. The same two-location reduction as Theorem 6 yields the claim with D replaced by d. A known vertex partition induces exactly such a disjoint partition of the edge coordinates into block pairs, proving Corollary 8. □

D.3 Proof of Lemma 9

Let $M = \{i : \widehat{z}_i \ne z_i\}$ after the optimal relabeling, so $|M| \le \gamma n$. Write the true block jump as $\Delta_{ij} = \Theta_{z_i z_j}$ and construct $\widetilde{\Delta}_{ij} = \Theta_{\widehat{z}_i \widehat{z}_j}$. Then $\widetilde{\Delta}$ lies in the estimated block-constant subspace. Its

orthogonal projector is $P_{\widehat{z}} = F_{\widehat{z}}^\top F_{\widehat{z}}$, hence

$$\|\Delta - P_{\widehat{z}}\Delta\|^2 \le \|\Delta - \widetilde{\Delta}\|^2.$$

The two matrices agree on pairs with both endpoints outside M. At most γn^2 undirected pairs touch M, and each discrepancy has absolute value at most $2\Delta_{\max}$, and we obtain $\|\Delta - P_{\widehat{z}}\Delta\|^2 \le 4\gamma n^2 \Delta_{\max}^2$. Since $\|\Delta\|^2 = D\Delta_{\mathrm{rms}}^2 = n(n-1)\Delta_{\mathrm{rms}}^2/2$, division proves (10). Finally, $\|F_{\widehat{z}}\Delta\|^2 = \|P_{\widehat{z}}\Delta\|^2$ because the rows of $F_{\widehat{z}}$ are orthonormal. □

D.4 A COMMON SCALAR FLOOR

Any structural family containing $\{p(\mathbf{1}\mathbf{1}^\top - I) : p \in [\rho/2, \rho]\}$ inherits a scalar lower bound from within that subfamily. Choose $p = \rho/2$ and $q = p + \delta/\sqrt{2D} \le \rho$. For one snapshot the divergence is at most

$$\sum_{e=1}^{D} \frac{(p-q)^2}{q(1-q)} \le 2\delta^2/\rho.$$

Two locations $u \asymp \min(T, \rho/\delta^2)$ apart therefore give the scalar floor when this radius is above a sufficiently large constant. The specified family must actually contain this witness; exact ranks greater than one or fixed unequal-degree constraints may exclude it.

D.5 PROOF OF PROPOSITION 10

Write $h = \theta - \beta$, $v = (\theta + \beta)/2$, $S = \mathbf{1}^\top v$. The difference of the completed matrices is $vh^\top + hv^\top$, with squared Frobenius norm

$$2\|v\|^2\|h\|^2 + 2(v^\top h)^2 \asymp n\rho\|h\|^2.$$

The removed diagonal has squared norm at most $4b^2\rho\|h\|^2$. For n sufficiently large this preserves the displayed order. The difference of the degree vectors is

$$\{SI + v\mathbf{1}^\top - 2\,\mathrm{diag}(v)\}h.$$

The inverse of $SI + v\mathbf{1}^\top$ equals $S^{-1}[I - v\mathbf{1}^\top/(2S)]$. Its operator norm is at most $C/(n\sqrt{\rho})$, while the forward norm is at most $Cn\sqrt{\rho}$. The diagonal perturbation has norm at most $2b\sqrt{\rho}$ and hence a vanishing relative size. We see that the degree difference has norm comparable to $n\sqrt{\rho}\,\|h\|$, and combining the comparisons proves the result. □

E FROZEN-REPRESENTATION AND SPLIT-TRANSFER PROOFS

E.1 PROOF OF THEOREM 11

Proposition 1 uses only independence across time and the two mean levels, so its population separation applies verbatim to $X_t = \psi(Y_t)$. Put $\epsilon_t = X_t - \mu_{j(t)}$, $u = |k - \tau|$ and $B = A_k - A_\tau$. The centered increment is

$$2\sum_t (Bz)_t\langle\mu_2 - \mu_1, \epsilon_t\rangle + \sum_{s\ne t} B_{st}\langle\epsilon_s, \epsilon_t\rangle.$$

For the linear term, scalar Bernstein together with $\mathrm{Var}\langle\mu_2 - \mu_1, \epsilon_t\rangle \le s_\psi\kappa_\psi^2$ and $|\langle\mu_2 - \mu_1, \epsilon_t\rangle| \le B_\psi\kappa_\psi$ gives

$$C_\eta\{\kappa_\psi\sqrt{s_\psi u x} + B_\psi\kappa_\psi x\}.$$

For the quadratic term, apply the bounded canonical order-two U-statistic inequality of Houdré & Reynaud-Bouret (2003) to the kernels $h_{st}(x, y) = B_{st}\langle x, y\rangle$. Lemma 16 and the covariance bounds give its four parameters, up to C_η, as

$$v_\psi\sqrt{u/T}, \qquad s_\psi\sqrt{u/T}, \qquad B_\psi\sqrt{s_\psi/T}, \qquad B_\psi^2/T.$$

We conclude that, outside an event of probability Ce^{-x}, the quadratic term is at most

$$C_\eta\left\{v_\psi\sqrt{ux/T} + s_\psi x\sqrt{u/T} + B_\psi\sqrt{s_\psi/T}\,x^{3/2} + B_\psi^2 x^2/T\right\}.$$

We union-bound over the scan and balance each term against the deterministic loss $c_\eta u\kappa_\psi^2$. The two square-root-in-u terms yield $(v_\psi^2 x + s_\psi^2 x^2)/(T\kappa_\psi^4)$; the remaining terms enter linearly, which proves (11). □

E.2 Proof of Proposition 12

Conditional on all even snapshots, $\widehat{\psi}$ is deterministic and the odd snapshots remain independent. Their first regime contains exactly $\tau_{\mathcal{O}} = \lceil \tau/2 \rceil$ odd indices. Since $\tau \in \mathcal{K}_\eta$ and $\eta T \geq 4$, this location belongs to the $\eta/2$-trimmed odd scan. We apply the relevant fixed-representation theorem (Theorem 2 or Theorem 11) conditionally with $T_{\mathcal{O}} = \lceil T/2 \rceil$ and trimming constant $\eta/2$. If it gives $|\widehat{\tau}_{\mathcal{O}} - \tau_{\mathcal{O}}| \leq R$, then

$$|2\widehat{\tau}_{\mathcal{O}} - \tau| \leq 2|\widehat{\tau}_{\mathcal{O}} - \tau_{\mathcal{O}}| + |2\lceil \tau/2 \rceil - \tau| \leq 2R + 1.$$

Averaging the conditional probability statement over the even sample, we obtain the unconditional claim. □

Finite-library corollary. Conditional on the even data, all M candidates are fixed. We apply the relevant odd-sample bound to candidate m with tail parameter $x + \log M$. Its failure probability is at most $CT_{\mathcal{O}} e^{-x}/M$. A union bound gives simultaneous validity for all candidates with total failure probability at most $CT_{\mathcal{O}} e^{-x}$, after which any even-measurable choice $\widehat{m}$ may be substituted. The argument holds for any selection rule. □

F How to read the minimax comparison and retention analysis

We interpret the lower bound as a statement about the cost of an unknown change direction. In the unrestricted class, an estimator must work simultaneously over many signed edgewise directions. The mixture construction in Theorem 6 makes those directions hard to distinguish and produces the $D\rho^2/(T\delta^4)$ term. If side information rules out most of those directions before the data are analyzed, the relevant difficulty can be much smaller. Corollary 8 makes that reduction explicit for known groups.

From this perspective we also see that a low-dimensional summary must be judged by the signal it retains as well as by its output dimension. A map with small rank may eliminate many nuisance directions, but it is useful for localization only if the true jump still has substantial image under that map. Conversely, a representation can have more coordinates than a competing summary and still be preferable because it preserves much more of the change. Our upper bound records this through the pair (d_F, κ_F) rather than d_F alone. In practice, we therefore judge a supplied aggregation by the changes it is capable of representing, not by its nominal dimension. The leverage ℓ adds a second qualification: two representations with the same rank and retained signal can expose very different amounts of a single Bernoulli edge to their feature space. For the nonnegative averaging maps used in the structural results, this leverage correction is lower order under the stated aggregation condition.

We see that the unrestricted and structured results are consistent. The unrestricted lower bound describes the price of robustness to an unknown direction; the structured lower bound describes a smaller experiment in which valid external information has already restricted that direction. The statistical question we study in the main text is precisely whether a particular supplied map reduces the former uncertainty without discarding too much of the jump.

F.1 A representation-level account of the gain

Let $s_0 = \|\Delta\|^2$ and let $a_F = \|F\Delta\|^2 / s_0 \in [0, 1]$ denote the retained fraction of edge-signal energy. The formulas below concern $a_F > 0$; if $a_F = 0$, the representation retains no change and its localization bound is vacuous. Keeping the tail parameter visible and assuming the aggregation conditions of Corollary 3, the relevant upper-rate expression is, up to constants,

$$R_F(x) = \frac{\rho x}{a_F s_0} + \frac{\rho^2 x(d_F + x)}{T a_F^2 s_0^2}. \tag{18}$$

For raw edges replace (a_F, d_F) by $(1, D)$. We see that the scalar summand is multiplied by $1/a_F$, while the quadratic summand is multiplied by $(d_F + x)/\{(D + x)a_F^2\}$. Compression improves the quadratic summand when

$$a_F > \sqrt{\frac{d_F + x}{D + x}},$$

which reduces to the familiar $\sqrt{d_F/D}$ heuristic only when the ranks dominate x. In particular, for $d_F \lesssim x \asymp \log(eT)$, the x^2 term creates a stochastic effective-dimension floor: reducing rank below order x no longer yields a proportional quadratic-noise gain. Since every lossy map has $1/a_F > 1$, an overall rate improvement requires more than a quadratic gain: the raw problem must be sufficiently dimension-dominated for the quadratic gain to offset the scalar efficiency loss. For known blocks in the balanced regime of Corollary 8, the dimension-dependent improvement is minimax sharp up to logarithms.

This comparison concerns the upper bounds, and our message is quantitative rather than categorical. A compressed representation can improve the dimension-dependent part of the rate only if it retains enough of the jump, and it can still lose overall when the raw problem is already dominated by the scalar term. The phase diagram itself conditions on a fixed representation; Section 5 shows how validity is preserved when a representation is learned or selected on an order-preserving holdout. Table 1 shows how the canonical fixed maps instantiate this phase diagram; its signal-behavior and consequence columns describe the population mean signal of each representation, independent of any data-driven selection.

F.2 Sufficient improvement regions for fixed structural compression

From the rate proxy in (18) we obtain a useful three-regime interpretation. First, if $a_F \approx 1$ and $d_F \ll D$, structure is nearly signal-preserving and primarily removes the direction-search penalty. Known balanced blocks are the cleanest example: $a_F = 1$ on the block model, so the scalar term is unchanged while the dimension term falls from D to d_{block}, and Corollary 8 shows that this improvement is minimax-sharp up to logarithms. Second, if $0 < a_F < 1$, compression trades a larger scalar term against a smaller quadratic term. Third, if $a_F = 0$—for example, a degree-preserving change analyzed only through degrees—the representation is blind to the mean change and no localization rate is available from that feature. We make the middle regime explicit with the proxy alone.

Within the proxy $R_F(x)$, define $G_F = (D + x) - (d_F + x)/a_F^2$. If $G_F \le 0$, the quadratic-noise reduction is too small to compensate for signal loss, so the proxy cannot beat raw edges. If $G_F > 0$ and $a_F < 1$, simple algebra gives

$$R_F(x) < R_{\mathrm{edge}}(x) \quad \Longleftrightarrow \quad s_0 < \frac{\rho}{T}\frac{G_F}{a_F^{-1} - 1}. \tag{19}$$

Lossy compression is most useful in the weak-signal, dimension-dominated regime. As the jump becomes strong, the scalar resolution term dominates and discarding signal can hurt even if the representation is much smaller. For $a_F = 1$ the scalar penalty vanishes, and any reduction of d_F improves the proxy's quadratic component. We use (19) as a sufficient upper-bound comparison.

Our fixed-map upper bound applies to supplied aggregation rules on an otherwise arbitrary Bernoulli network; the bound depends on whatever $\|F\Delta\|$ the map retains. The structural model enters only when we want a guarantee that the map must preserve signal (Proposition 10) or when we compare against a matching structured lower bound (Corollary 8). Corollary 5 also gives a practical route: communities or another linear summary may be learned on an independent pilot and then treated as fixed. Proposition 12 gives a same-sequence route via parity. This distinction separates our use of side information from methods that jointly estimate network-model parameters or communities (Bhattacharjee et al., 2020; Peel & Clauset, 2015).

F.3 Relation to Wang–Shao

Under homoscedastic representation noise with covariance Σ_F, the normalization of Wang & Shao (2023) reads, in our notation, $a_T = T^2\kappa^4/\|\Sigma_F\|_F^2$. We see that their integer-scale error T/a_T is exactly $\|\Sigma_F\|_F^2/(T\kappa^4)$, the analogue of the second term in (7), and their break-dating rate $\|\Sigma\|_F^2/(T\|\Delta\|^4)$ for the distinct-index U-statistic reduces to our $d\rho^2/(T\kappa^4)$ scale under nearly homoscedastic noise. We read our upper result as a finite-sample, Bernoulli-heteroscedastic, representation-aware version of that phenomenon; our new information-theoretic point is that the corresponding dimension term is necessary in general; Theorem 6 proves this directly for Bernoulli mean changes over the full window $\rho\sqrt{D} \lesssim T\delta^2 \lesssim \rho D$. We use the corrected U-type contrast as

the vehicle to attain the matching upper rates; its distinct-index construction is closely related to Wang–Shao, and our main novelty lies in the lower bounds of Theorems 6–7.

F.4 Expanded positioning

High-dimensional change-point analysis. High-dimensional localization work has largely emphasized first-order signal-to-noise resolution or sparse projections; see the optimal detection and localization analysis of Verzelen et al. (2023) for univariate mean changes, as well as Wang & Samworth (2018), Kaul et al. (2021), and the minimax overview of Yu (2020). Dimension-dependent terms also appear prominently in detection boundaries (Liu et al., 2021; Pilliat et al., 2023). Recent covariance-scanning work in high-dimensional regression develops related lower-bound questions for detection (Cho & Li, 2025). We cite its June 2026 v4, whose stated lower-bound claim concerns detection in a different model. We also distinguish our setting from common-break mean/variance work such as Bai (2010): the Bernoulli model couples the mean and variance automatically, so ordinary squared CUSUM can drift even though the target is a mean change.

Dynamic-network change points. Dynamic-network change-point methods span model-based, matrix, spectral, graph/kernel, and learned comparisons. Early work includes graph-based scans (Chen & Zhang, 2015), large-scale structural comparison (Peel & Clauset, 2015), and sequence-of-graph methods (Zambon et al., 2019). For stochastic network models, Bhattacharjee et al. (2020) study a dynamic stochastic block model and Wang et al. (2021) give optimal detection and localization results for sparse independent Bernoulli networks; related formulations use dependent random-dot-product graphs (Madrid Padilla et al., 2022), Laplacian spectral summaries (Huang et al., 2020), learned graph similarities (Sulem et al., 2024), and latent-space decoders (Kei et al., 2025). These methods are natural empirical comparators that answer different questions about model estimation, graph geometry, or learned similarity. We keep the network model deliberately simple so that we can isolate an information-theoretic question about a supplied representation: which jump directions it retains, which unresolved directions it removes, and how those two effects alter localization difficulty.

Quadratic contrast and variance drift. The statistic we analyze is a distinct-index U-type contrast closely connected to Chen & Qin (2010) and, more directly, to the break-dating objective of Wang & Shao (2023). Its population dependence on means rather than variances follows from that distinct-index construction; Proposition 1 records the identity in our notation. In the Bernoulli network setting, we see a concrete failure mode of the ordinary squared CUSUM: because mean and variance move together, its population maximizer can leave the true split. The U-type form avoids this drift and is amenable to variance-sensitive concentration (Giné et al., 2000).

F.5 Full scope conditions and limitations

We treat one change and independent snapshots. The parity result supplies held-out validity for a prespecified finite library; it is not an oracle-optimal same-data selector, an adaptive signal certificate, or a recovery of the efficiency lost to splitting, and it neither identifies a statistically optimal selection score nor proves an oracle inequality. We reserve those stronger adaptive questions for future work. The nonlinear theorem assumes bounded frozen outputs and is a conservative transfer result rather than a sharp general theory for learned GNN embeddings; sharper unbounded and representation-specific results are left for future treatment, and because the bounded-range terms can dominate for some maps, the linear Bernoulli theorem remains the sharper result when edge-level structure is available. The lower bound for structural compression is proved for aligned disjoint aggregations; a general non-aligned lower bound would be needed to turn the sufficient crossover proxy into a rate characterization for arbitrary maps; such a bound would require controlling what information remains in the distribution of FY_t even when $F\Delta$ is small. The unrestricted class of Theorem 6 is dense in the jump direction, and the same second term is not claimed for an s-sparse jump, where an s-dependent analogue is a natural open question. The worst-case leverage term of Theorem 2 is not claimed to be necessary in the sparse regime $\rho < \ell x$. Lemma 9 is a worst-case vertex-misclustering bound and can be loose when the misclustered vertices carry little jump mass. Proposition 10 has no matching n-term lower bound over the whole Chung–Lu class, and the observed zero-diagonal matrix is not literally rank one. Multiple changes, temporal dependence, variance-adaptive certification, and formal uncertainty quantification for the Enron sequence are outside the present scope. The

constant-mean scan of Appendix B establishes null control, not consistency of a multiple-change selection rule, and neither exchangeability nor its randomization argument holds automatically under temporal dependence. The fixed-map comparison of Appendix F.1 concerns upper bounds and makes no claim that every fixed representation attains an exact minimax rate on every parameter class. For Enron, the whole-snapshot permutation diagnostics do not identify a change date, and neither they nor the fitted-regime ranges are used as theorem-backed inference.

G Focused experiment details

G.1 Synthetic data-generating mechanisms

We use Bernoulli draws independent over time and edges in every synthetic configuration. The primary signal/retention experiments use $n = 96$, $T = 160$, $\tau = 64$, baseline edge probability 0.2, and a 20% trim. Four consecutive groups of 24 vertices define the supplied partition.

For the signal sweep, the block direction assigns one fixed sign to each of the ten unordered block pairs and is renormalized at each $\delta^2 \in \{0.1, 0.2, 0.4, 0.6, 1, 1.8, 3\}$. The Chung–Lu family starts from $\theta_i = \sqrt{0.2}$ and perturbs half of the vertex parameters upward and half downward; bisection sets the full Frobenius jump. The degree-blind direction is constructed to have exactly zero row sums. Every cell uses 200 replicates and common sequences across methods.

For the retained-fraction sweep, $u_{\|}$ is a unit vector in the row space of the normalized four-block map and $u_{\perp}$ is a unit vector in its null space. We use $a_F \in \{0, .05, .10, .20, .40, .60, .80, 1\}$ with

$$\Delta(a) = \sqrt{\delta^2/2}\{\sqrt{a}\, u_{\|} + \sqrt{1-a}\, u_{\perp}\}.$$

Because the block map has orthonormal rows, the realized retained fraction equals a up to floating-point error. The original E1 sweep uses $\delta^2 \in \{0.6, 1.8\}$; the E4 stress test uses $\delta^2 \in \{3, 5, 8\}$. Each cell has 200 replicates. E4 additionally records paired Block-minus-Edge absolute-error differences and their Monte Carlo standard errors. The vertical proxy markers use the unit-constant expression $R_F(x)$ with $x = \log(eT)$, $T = 160$, $\rho = 0.2$, $D = 4560$, $d_F = 10$, and $s_0 = \delta^2/2$. The E1 marker values are 0.088 and 0.156; the E4 values are 0.221, 0.312, and 0.417 for $\delta^2 = 3, 5, 8$. These are reproducible sufficient-proxy values computed with unit constants.

For sparse scaling, $n = 64$, $T = 240$, $\tau = 96$, and $p_{1,e} = 0.75\rho$. Fixed random edge signs define $p_{2,e} = p_{1,e} + hs_e$ with $h^2 = \delta^2/(2D)$ and $\delta^2 = \lambda\rho\sqrt{D}/T$. We use $\rho \in \{0.04, 0.08, 0.16, 0.32\}$ and $\lambda \in \{1, 2, 4, 8, 16, 32\}$ with 200 replicates per configuration.

G.2 Numerical detail for the main figures

Signal sweeps (Figure 1). At $\delta^2 = 0.6$, Edge U has MAE 24.42, 25.69, and 24.28 in the block, Chung–Lu, and degree-blind families, respectively, whereas the aligned Block U, Degree U, and Block U errors are 1.85, 4.10, and 1.82. When the full jump reaches $\delta^2 = 3$, we see the unstructured statistic recover to MAE 6.09, 5.44, and 5.21; at the same strong signal the aligned representations have MAE 0.13, 0.21, and 0.25. Degree U remains at MAE 32.85 in the degree-blind family because its mean change is exactly zero. Edge U and Wang–Shao remain numerically nearly indistinguishable, as expected from their algebraic relationship.

Moderate-signal retention (Figure 2). The displayed proxy values 0.088 and 0.156 are obtained by inserting $x = \log(eT)$, $T = 160$, $\rho = 0.2$, $D = 4560$, $d_F = 10$, and $s_0 = \delta^2/2$ into the unit-constant proxy (18), with unit constants in place of the unspecified theorem constants. At $\delta^2 = 0.6$, Block U and Edge U are essentially tied at $a_F = 0.05$ (MAE 25.66 versus 25.70), while Block U is better at $a_F = 0.10$. At $\delta^2 = 1.8$, the empirical advantage also begins between 0.05 and 0.10, earlier than the proxy value 0.156. We see that the moderate-signal sweep supports the qualitative retention mechanism; the strong-signal extension below tests the opposite regime, in which a strong uncompressed signal should make an overly lossy structural summary worse.

Strong-signal retention (Figure 3). At $a_F = 0$ the block map erases the jump and performs near the null regime: Block U has MAE 31.50, 31.88, and 31.05 for $\delta^2 = 3, 5, 8$, while Edge U has MAE 5.44, 1.92, and 0.50. Even $a_F = 0.05$ remains harmful: the paired mean difference

$\mathrm{MAE}_{\mathrm{Block}} - \mathrm{MAE}_{\mathrm{Edge}}$ is 6.06, 5.31, and 2.40. The sign reverses as enough signal survives. For $\delta^2 = 3$, Block U is essentially tied with Edge U at $a_F = 0.10$ and clearly better from $a_F = 0.20$ onward; for $\delta^2 = 5$, the paired difference is still positive at 0.10 but negative at 0.20; for $\delta^2 = 8$, the mean difference is slightly negative at 0.20 but its paired 95% Monte Carlo interval still includes zero, while $a_F = 0.40$ clearly favors Block U. The corresponding unit-constant sufficient-proxy crossings are 0.221, 0.312, and 0.417. We see the observed sign changes occur before these crossings, as one would expect from a bound with unspecified constants; the proxy serves only as a sufficient upper-bound comparison. Our results therefore support the theorem's direction: strong raw signal makes lossy compression observably harmful, and sufficient retention reverses that ordering.

Independent pilot (Figure 4). The figure shows the oracle partition and the fixed-midpoint MAE for reference. At $\delta^2 = 0.6$, increasing T_0 from 5 to 40 reduces mean misclustering from 0.586 to 0.063, raises the realized retained fraction from 0.042 to 0.789, and reduces Pilot Block U MAE from 24.50 to 1.95; the oracle Block U MAE at $T_0 = 40$ is 1.34. At $\delta^2 = 1.8$, the corresponding retained fraction rises from 0.056 to 0.718 and Pilot Block U MAE falls from 13.55 to 0.60, versus oracle MAE 0.36. Our controlled-misclustering experiment verifies the direction of Lemma 9: corrupting about 5%, 10%, and 20% of labels yields mean retained fractions 0.77, 0.59, and 0.34, while the deterministic lemma gives conservative lower bounds 0.58, 0.16, and 0. At $T_0 = 20$, we find that mild misspecification of the number of blocks is also stable in this design: for $\delta^2 = 0.6$, Pilot Block U has MAE 7.53, 9.34, and 8.34 for $K \in \{3, 4, 5\}$, with mean retained fractions 0.253, 0.317, and 0.298; for $\delta^2 = 1.8$ the corresponding MAEs are 1.36, 1.18, and 1.21, with retained fractions 0.389, 0.407, and 0.382.

G.3 Independent-pilot partition experiment

Our pilot experiment uses a four-block SBM whose partition is common before and after the change. The pre-change block matrix has probability 0.18 within groups and 0.14 between groups. A fixed symmetric block-jump pattern is scaled to $\delta^2 \in \{0.6, 1.8\}$. The independent pilot sequence contains half pre-change and half post-change graphs; because both regimes share the partition, a straddling pilot retains the structural side information while remaining independent of the localization sequence.

For $T_0 \in \{5, 10, 20, 40\}$, we apply spectral clustering to the mean pilot adjacency matrix after removing its global-density component. We report the optimal-label misclustering fraction, the realized retained fraction of the resulting normalized block map, and localization error on a fresh analysis sequence. We include misspecified $K \in \{3, 5\}$ variants at $T_0 = 20$ in the released tables. A separate stress test randomly corrupts approximately 5%, 10%, or 20% of the true labels, directly checking the deterministic inequality in Lemma 9. All cells use 200 replicates.

G.4 Parity-split finite-library experiment

E5 uses the same $n = 96$, $T = 160$, $\tau = 64$, and 20% trim. Its library contains raw edges, normalized degrees, and block maps obtained by spectral clustering with $K \in \{2, 3, 4, 5, 6\}$. For block-based candidates the partition is estimated from the selection sample, so the library is data-dependent but measurable with respect to that sample. We release full candidate-by-candidate selection frequencies with the repository.

For a selection sample of length T_s, the operational score is exactly (12): $x_s = \log(eT_s)$, the raw plug-in is $T_s U_c^{\max} / \{\widehat{k}_c(T_s - \widehat{k}_c)\}$ truncated below at zero, the floor is $0.30\sqrt{(d_c + x_s)/T_s}$, and the two-term proxy uses the same global scoring constant 0.30. The final E5 script fixes this value before the simulation loop. The Round-4 sensitivity notebook reruns the selector with the floor alone multiplied by 0.5, 1, and 2, while leaving the proxy constant and candidate library unchanged.

The parity arm learns/selects on even snapshots and localizes on odd snapshots, returning $2\widehat{\tau}_{\mathcal{O}}$. The independent-pilot arm uses a separate length-80 sequence, with its change at 32, so its selection budget equals one parity half; after selection it localizes on all 160 analysis snapshots. The naive arm selects and localizes on the same 160 observations, and we report it descriptively. The oracle uses normalized degrees for Chung–Lu and the true four-block map for block and degree-blind changes. Each of the three designs and $\delta^2 \in \{0.6, 1.8, 3\}$ has 200 replicates.

Floor sensitivity. The prespecified E5B diagnostic multiplies only the floor term f_c by $\{0.5, 1, 2\}$; the global proxy constant 0.30, candidate library, seeds, and localization rule are unchanged. We report parity-split MAE with Monte Carlo SE in Table 6. We see that the result is stable for block and degree-blind changes at moderate/strong signal, while Chung–Lu is sensitive to an aggressive floor, consistent with the lower selection frequency of the degree family in Table 2.

Table 6: E5B floor sensitivity over 200 paired replicates. Entries are parity-split MAE with Monte Carlo SE in parentheses. Only the detectability floor is multiplied; the selector proxy and library are otherwise unchanged.

Design	δ^2	$0.5\times$ floor	$1\times$ floor	$2\times$ floor
Block	0.6	4.76 (0.52)	4.75 (0.50)	9.12 (0.95)
Block	1.8	0.97 (0.15)	0.99 (0.15)	0.96 (0.14)
Block	3	0.53 (0.08)	0.54 (0.08)	0.56 (0.08)
Chung–Lu	0.6	27.60 (1.33)	27.12 (1.31)	27.75 (1.33)
Chung–Lu	1.8	6.05 (0.75)	8.94 (0.97)	22.50 (1.39)
Chung–Lu	3	2.29 (0.52)	2.03 (0.41)	7.00 (1.05)
Degree-blind	0.6	5.11 (0.56)	4.58 (0.45)	4.61 (0.51)
Degree-blind	1.8	0.54 (0.08)	0.54 (0.08)	0.54 (0.08)
Degree-blind	3	0.31 (0.06)	0.29 (0.06)	0.27 (0.06)

G.5 Portable fixed-representation comparators and canonical package audit

The original E2 comparison at $\delta^2 = 1.8$ uses four additional predetermined summaries. The RBF scan first applies a fixed 64-dimensional sparse CountSketch to every edge vector and uses the median pairwise squared distance as the RBF bandwidth. The MST scan builds a minimum spanning tree of the same sketch vectors and scores candidate splits by the standardized deficit of tree edges crossing the split, using 60 random-order null permutations per replicate. The frozen spectral map uses the six eigenvalues of largest absolute value from the normalized adjacency matrix and then applies corrected U. The Laplacian-signature scan compares short and long moving-window spectral signatures using cosine distance, with windows 6 and 12. We keep these as compressed-representation ablations; the package-identical baselines follow below.

For completeness, their E2 MAEs at $\delta^2 = 1.8$ are:

Design	Edge U	Wang–Shao	Aligned U	RBF sketch	MST sketch	Spectral U
Block	14.57	14.51	0.50	25.98	26.77	31.99
Chung–Lu	12.40	11.99	0.60	32.14	30.44	28.60
Degree-blind	13.54	13.61	0.55	28.10	29.13	31.81

We retain these rows as representation controls.

The frozen nonlinear controls remain useful for a different reason. A six-eigenvalue spectral representation and a short/long-window Laplacian signature can be highly compressed yet preserve little of the relevant mean change; at $\delta^2 = 1.8$ frozen Spectral U has MAE 28.60–31.99 where the aligned maps reach 0.50–0.60. We see that their poor performance is consistent with Theorem 11, which is explicitly controlled by κ_ψ and covariance effective rank; it is another instance of our main distinction between low nominal dimension and retained change information.

E6 separately evaluates raw-input implementations in R 4.4.3: `gSeg` 1.1, `kerSeg` 1.1, `changepoints` 1.1.0, `oddnet` 0.1.2, `ade4` 1.7.24, and `igraph` 2.3.1. For gSeg we build the MST from Euclidean distances between raw edge vectors and use the max-type single-change statistic. For kerSeg we use the Gaussian kernel on raw edge vectors with the median pairwise squared distance as the bandwidth scale. Network WBS uses $M = 100$ random intervals in final mode and `delta=5`. The input is the strict upper triangle, so the package matrix is $D \times T = 4560 \times 160$; odd and even columns form the two 4560×80 copies with `self=FALSE`. For Table 3 we use the finite candidate with largest `Dval`, not `thresholdBS`, so there is no WBS detection threshold and we map every replicate to one location by doubling the half-grid index. For local refinement, $\widehat{\rho}$ is the 0.95 quantile of the edgewise full-sequence means, $w = 0.5$, $\tau_2 = 96\widehat{\rho}/3$, and $\tau_3 = \infty$; the refined

half-grid location is again doubled. LAD uses raw adjacency matrices with $k = 20$, short window 6, and long window 12, selecting the maximum over the admissible scan. E6 uses 100 replicates for each design and $\delta^2 \in \{0.6, 1.8, 3\}$; all 4,500 underlying method-level records completed successfully, including the 3,600 records displayed in Table 3.

The prespecified E6B diagnostic uses 50 fresh replicates per cell at $\delta^2 \in \{8, 20\}$. It evaluates the package-example WBS threshold $\tau_{\mathrm{WBS}} = 96\widehat{\rho}\{\log(80)\}^2/20$, compares it with the strongest finite WBS candidate, initializes refinement both at WBS and at the true half-grid change, sweeps τ_2 over $\{0.5, 1, 2\}$ times the reported default with $\tau_3 = \widehat{\rho}$, and replaces the single MST by a denser 5-NN similarity graph for the gSeg sanity check. We report the most diagnostic subset in Table 7 (Table 4 in the main text shows its core columns); WBS-initialized refinement gives the same qualitative threshold-sensitivity conclusion as oracle initialization. All 3,600 E6B method-level records completed successfully.

Local refinement is different. Its originally reported thresholding remains poor in several strong-signal cells even when initialized at the true change point, whereas halving τ_2 can improve the result dramatically. For example, with oracle initialization at $\delta^2 = 8$, the reported setting gives MAE $(16.68, 30.00, 26.92)$ across the three designs, while the $0.5\tau_2$ setting gives $(5.20, 6.32, 6.44)$. We see that the anomaly arises from the refinement threshold rather than from WBS initialization, and that the refinement conclusion is highly threshold-dependent in these settings. Rather than select a favorable threshold after observing the diagnostic, we remove local refinement from the headline comparison and report the full sanity check in Table 7.

Table 7: E6B strong-signal package sanity check over 50 replicates per cell. Entries are MAE (Monte Carlo SE). "Refine default" uses the true change as initializer with the originally reported τ_2 and $\tau_3 = \infty$; "Refine $0.5\tau_2$" uses the true initializer, half the default τ_2, and $\tau_3 = \widehat{\rho}$.

Design	δ^2	WBS max	WBS threshold	Refine default	Refine $0.5\tau_2$	gSeg MST	gSeg 5-NN
Block	8	1.00 (0.36)	1.00 (0.36)	16.68 (2.01)	5.20 (0.98)	19.68 (3.00)	6.24 (1.43)
Block	20	0.00 (0.00)	0.00 (0.00)	0.08 (0.06)	0.40 (0.16)	13.56 (2.51)	7.04 (1.76)
Chung–Lu	8	2.04 (0.53)	2.04 (0.53)	30.00 (0.00)	6.32 (1.35)	20.40 (3.20)	9.08 (1.91)
Chung–Lu	20	0.04 (0.04)	0.04 (0.04)	19.32 (1.96)	0.44 (0.17)	7.36 (1.88)	1.10 (0.23)
Degree-blind	8	1.60 (0.47)	1.60 (0.47)	26.92 (1.20)	6.44 (1.10)	17.34 (2.74)	5.20 (1.00)
Degree-blind	20	0.04 (0.04)	0.04 (0.04)	0.16 (0.08)	0.40 (0.18)	5.46 (1.72)	1.00 (0.23)

G.6 NULL-ARGMAX DIAGNOSTIC

E7 uses 300 replicates per cell at $\delta^2 = 1.8$. The degree-blind change is paired with a no-change Degree U control; the aligned block change is passed through the six-eigenvalue frozen spectral map and paired with a no-change Spectral U control. For each replicate we record the maximizing split, absolute error relative to $\tau = 64$, and whether the maximizer falls in the outer quartiles of the admissible scan. Our purpose is diagnostic: if a representation carries negligible usable change signal, its change and no-change argmax distributions should look similar. Table 8 summarizes the four cells and Figure 5 shows the histograms.

Table 8: Null-argmax diagnostic (300 replicates, $\delta^2 = 1.8$). Boundary mass is the fraction of estimates falling in the outer quartiles of the admissible scan.

Cell	MAE	Median AE	Boundary mass
Degree U, degree-blind change	32.19	31	0.663
Degree U, no change	30.49	28	0.610
Spectral U, block change	30.62	27	0.630
Spectral U, no change	31.65	30	0.613

The change and no-change rows are strikingly similar, and all put more than 60% of their mass in the outer quartiles of the scan. We see that the below-uniform MAEs of these deliberately uninformative representations reflect a null-like boundary-concentration phenomenon of the argmax, distinct from a statistic that tracks a wrong interior change; we use the comparison purely as a diagnostic.

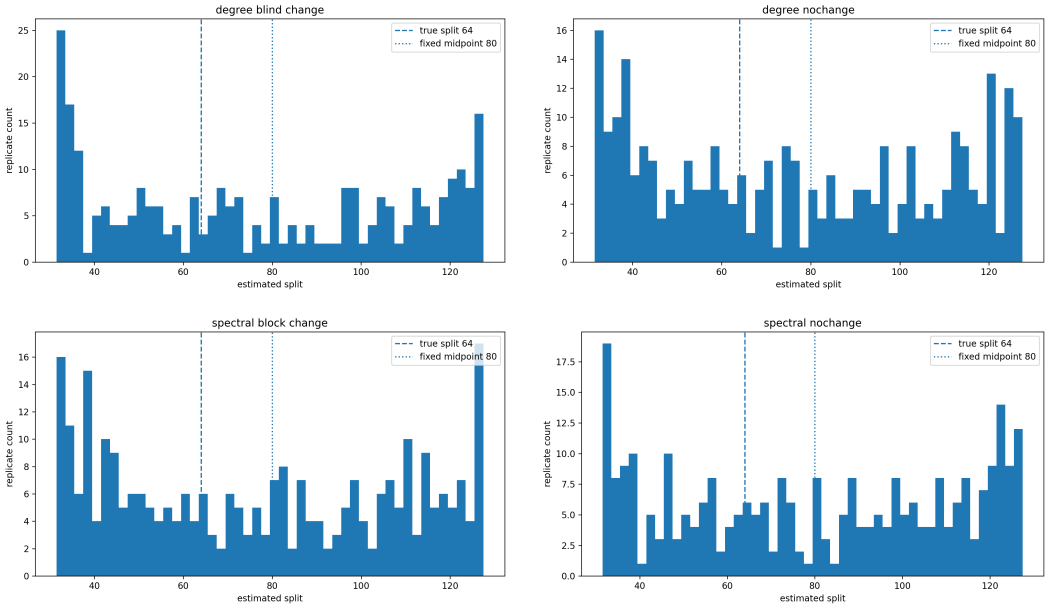


Figure 5: Null-argmax diagnostics. Top: Degree U under a degree-blind change and under no change. Bottom: frozen Spectral U under a block change and under no change. In both pairs the distributions are similar and place substantial mass near the scan boundaries.

G.7 Finite-sample variance drift and sparse scaling

Our retained-signal experiments use the corrected contrast throughout, and two diagnostics isolate why that correction matters. Figure 6 shows the exact population defect of raw squared CUSUM together with the sparse-scaling behavior of Edge U (design in Appendix G.1), while Figure 7 shows the finite-sample drift toward the wrong population maximizer of Proposition 14. Table 9 reports a finite-sample drift example.

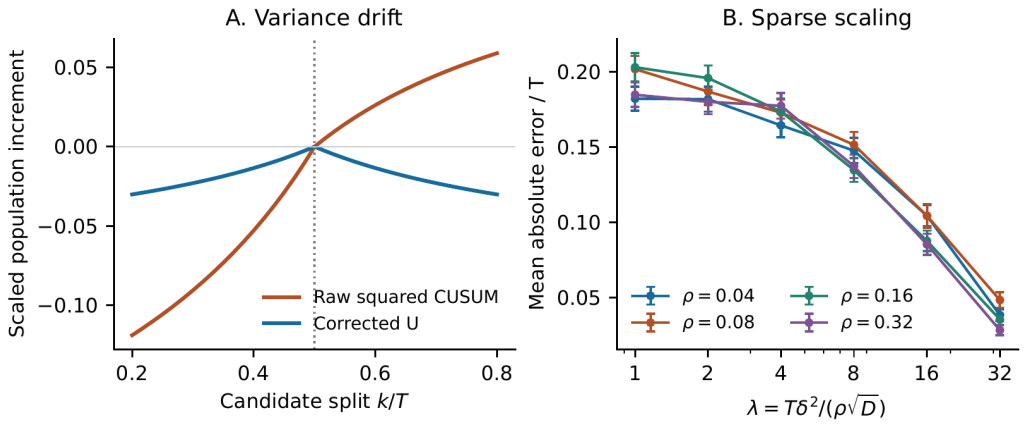


Figure 6: Left: exact population increments $T\{\mathbb{E}S_k - \mathbb{E}S_\tau\}/D$ for the raw squared CUSUM Q and corrected contrast U ($T = 60$). Right: Edge U localization over four sparsity levels with signed edgewise jumps, $n = 64$, $T = 240$, $\tau = 96$.

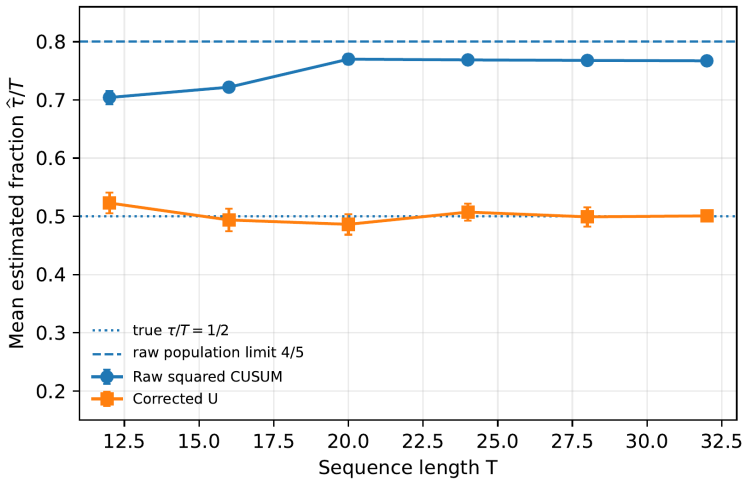


Figure 7: Finite-sample visualization of Proposition 14 (40 replicates per T). Corrected U remains at the true fraction $1/2$, while the raw squared CUSUM approaches its incorrect population maximizer $4/5$.

Table 9: Drift example: MAE $\pm$ Monte Carlo standard error (100 replicates).

n	Raw edges	Edge U	Wang–Shao
120	10.94 ± 0.57	10.60 ± 0.55	10.60 ± 0.55
240	12.48 ± 0.52	9.82 ± 0.56	9.55 ± 0.55
480	13.09 ± 0.50	8.92 ± 0.63	8.63 ± 0.63

The finite table illustrates finite-sample behavior; the exact-localization statement of Proposition 14 is asymptotic.

G.8 Enron construction and robustness protocol

The public Benson/Cornell archive contains timestamped email simplices among 143 core addresses. We retain messages with at least two distinct participants, project each message to all undirected pairs, and mark an edge active on a day if its pair co-occurs in at least one email that day. We decode the archive's timestamp field as ordinal-day milliseconds.

The primary E3 window is July 1–November 15, 2001 with trim 0.15. We analyze all daily snapshots, weekday-only daily snapshots, and Monday-start weekly binary aggregations. The edge feature has dimension $10{,}153$; the normalized degree feature has dimension 143. We perform whole-snapshot permutations within day-of-week strata for the daily constructions and at the whole-week level for weekly aggregation. The fitted-regime resampling scheme draws separately from the two estimated sides while preserving day-of-week strata for daily data. We report its 2.5–97.5 percentile location range as a descriptive stability summary. The whole-snapshot permutation uses 999 random orders in final mode, so the minimum attainable reported value is $1/1000 = 0.001$. Because the observed sequence is dependent, we treat both diagnostics as descriptive.

Our sensitivity analysis varies the window start and end independently by ± 14 days and uses trims $\{0.05, 0.10, 0.15, 0.20\}$, producing 36 configurations for each calendar/representation pair. We rerun the earlier July 1–December 31, trim-.20 analysis as a reproduction audit and exactly recover September 22 for edges and September 25 for degrees.

Weekly estimates are necessarily coarser: across the sensitivity grid they lie in the weeks beginning September 17 or September 24, with median September 24 for both representations. A Monday-start week beginning September 24 contains September 26, so weekly aggregation can mix observations around a daily transition. The primary weekly resampling ranges are centered on the week-of-September-17 fit and therefore stop one bin earlier, whereas the window/trim multiverse includes

configurations whose maximizer shifts to the week of September 24; this explains why the multiverse median can be later than the primary resampling range. We interpret the weekly analysis at weekly resolution. The whole-snapshot permutation diagnostics test whether a no-change ordering is compatible with the observed maximum, and the fitted-regime resampling ranges are likewise descriptive. Because the daily and weekday-only multiverse directly targets the weekend concern, we report it as part of the main result (Section 6.4). An independent latent-space analysis reports a nearby late-September change (Kei et al., 2025), which we treat as contextual evidence.