Identifiability and Recovery of Episodic Coordination in Switching Marked Point Processes

Anonymous authors
Paper under double-blind review

Abstract

We study when episodic coordination is identifiable and recoverable in regime-switching marked Hawkes processes, where a brief coupled episode is easily confused with burstiness, time-averaged coupling, or segmentation artifacts. We show recovery requires four ingredients: sufficient effective active exposure, a source of dormant–active separation, birth-time attribution of excitation, and a normalization separating actor couplings from mark-transition weights. A finite-sample analysis identifies birth-time active-born events from true actors as the effective sample size and, under an explicit posterior-accuracy condition, transfers standard sparse support-recovery machinery to this setting, with weak links requiring quadratically more exposure. These conditions motivate a birth-time gated marked Hawkes estimator with role-separated sender/receiver gates, asymmetric low-rank directional structure, and generalized-EM activation recovery. Synthetic experiments, including an external change-point Hawkes baseline, internal identifiability ablations, sparse directed-edge diagnostics, and posterior-accuracy stress tests, validate the exposure threshold, support the posterior-accuracy bridge used by the recovery theorem, and show the predicted collapse when separation conditions are violated. A dedicated synthetic weakest-link sweep confirms the predicted difficulty ordering where weaker links require longer active exposure. Empirically, planted-coordination and predictive checks show that the estimator recovers member and receiver sets on volume-neutral UCDP GED conflict-event data, a cross-asset mark-transition channel on FTX-collapse trade-event data, and the same recovery ordering on real hippocampal CA1 recordings, while exposure, separation, or attribution failures lead to non-identifiability or practical recovery failure.

1 Introduction

Many scientific and operational systems are observed as streams of timestamped events rather than continuous trajectories: a neuron fires, an animal initiates a movement, an account accesses a file, a financial asset is traded. In each case the time, the actor, and the event type all matter, and marked point processes capture all three in one likelihood-based framework. The signal of coordination is rarely just an elevated rate; it is typically a structured transition from one event type to another—stimulus to spike, privilege escalation to file access, stress in one asset to large trades in another. The problem sharpens when coordination is *episodic*: a small subset of actors couples for a short window and then returns to routine behavior, as in a transient neural assembly, a brief behavioral synchronization, a market-stress cascade, or a short burst of collusive activity. Across these domains the statistical question is the same: from the event stream alone, can we recover *when* an episode occurred, *who* participated, and which actors acted as *senders* versus *receivers*?

Existing tools do not recover this structure, because the target is a coordination *shape*—a directed, marked, time-localized excitation operator—not an anomaly score, and a scalar score collapses exactly the structure (who influenced whom, through which mark transition, under which regime, at what birth time) that explains coordination. Time-invariant multivariate Hawkes models average a brief episode against long stretches of inactivity, blurring its timing and direction. Latent-state and Markov-modulated Hawkes models relax stationarity but evaluate the regime at the *current* time—severing events from the regime that produced them—and treat the active interaction matrix as an unstructured block rather than a sparse set of participants

with directional structure. None ties its guarantees to how much active-regime evidence the data actually contain. The deeper obstacle is identifiability, not estimation: several distinct mechanisms generate nearly the same stream—routine excitation mimics weak active coupling, time-averaged structure absorbs episodic bursts, an HMM segments busy intervals without assigning actors, and actor coupling trades off against mark sensitivity—so episodic coordination is first an identifiability problem and only second an estimation one.

We characterize this identifiability boundary and design an estimator around it. The active kernel block is identifiable only when the process enters the active state, active-born events supply nondegenerate exposure, and dormant and active excitation are separated—by birth-state information or, for exponential kernels, by distinct decay rates $\beta_0 \neq \beta_1$—together with a normalization of the mark weights, since $\alpha^{(1)}_{j\to i}\rho^{(1)}_{m'm}$ is otherwise invariant to reciprocal rescaling. Finite-sample recovery is governed by the *effective active exposure* N_{eff}, the number of active-born events generated by true coordinated actors; identifying N_{eff} as the operative sample size is the model-specific content of our recovery theorem, whose remaining machinery (CTMC occupation, counting-process concentration, restricted-eigenvalue support recovery, and M-estimator perturbation) is standard. The guarantee is explicitly conditional on posterior accuracy—the posterior-weighted E-step must stay uniformly close to the oracle birth-state posterior at the scale of Assumption 2—so Theorem 3 transfers oracle support recovery to the implemented estimator rather than proving posterior concentration from primitive emission and mixing conditions. The bound also shows weak links require quadratically more exposure while the candidate population enters only logarithmically. These conditions dictate the estimator: birth-time gating keeps each event's regime label attached to the excitation it causes; role-separated sender and receiver gates make participant recovery a sparse-support problem; asymmetric low-rank embeddings encode direction among the surviving actors; and mark-transition normalization removes the actor–mark scaling ambiguity.

Our contributions are:

1. We characterize identifiability failures in switching marked Hawkes processes, including kernel collinearity (Theorem 2), actor–mark scaling ambiguity (Proposition 1), birth-time attribution (Theorem 1, D1 ablation in §6.3), and modular segmentation without joint assignment.
2. We prove active-kernel identifiability under dormant–active separation (Theorem 1), identify the birth-time effective exposure N_{eff} as the relevant sample size for gate recovery, and derive a conditional finite-sample support-recovery result by applying standard sparse M-estimation machinery under oracle curvature and posterior-accuracy assumptions (Theorem 3).
3. We develop a birth-time gated marked Hawkes EM estimator that repairs these failures by jointly inferring the activation timeline, participant support, directional role structure, and mark-specific propagation channels (Table 1).
4. We validate the theory through experiments organized around four theorem-facing predictions: recovery should improve with realized active-born exposure (P1), weakest-link recovery should require quadratically more active exposure as the weakest participation strength decreases (P2), recovery should be conditional on posterior accuracy at the theorem scale (P3), and removing the separability repairs should produce the predicted collapse modes (P4).

Table 4 maps the empirical evidence onto these four predictions, distinguishing direct support from diagnostic or negative evidence: synthetic experiments give the cleanest known-truth tests, UCDP GED (Sundberg & Melander, 2013; Shawn Davies, 2026) a volume-neutral real social-science exposure test, CRCNS CA1 spike-ins a real neural background, raw CA1 likelihoods a label-free predictive diagnostic, and FTX a cross-domain marked-channel probe on a real market-stress background. The theoretical counterpart is Table 1.

2 Related Work

A large body of work estimates a single time-invariant triggering structure for multivariate Hawkes processes. Zhou et al. (2013b) estimate triggering kernels nonparametrically via majorization-minimization, Bacry et al. (2015) give a broad treatment in the financial setting, Xu et al. (2016) study Granger causality for Hawkes processes, and Bacry et al. (2020) combine sparse and low-rank regularization on the kernel matrix. Linderman & Adams (2014) infer latent network structure under a Bayesian prior, and Zhou et al. (2013a)

model social infectivity as sparse-plus-low-rank. These methods are the natural points of comparison for the active-regime block of our model, but they do not address when coupling is present. Wang et al. (2012) introduce a Markov-modulated Hawkes process with stepwise decay, and Wu et al. (2022) develop inference for sporadic and bursty social interaction streams under a two-state Markov-modulated Hawkes model. These works show that latent-state modulation can capture heterogeneous event dynamics, but they address a different level of structure than we do. In particular, Wu et al. (2022) focus on a single interaction stream, such as a directed pairwise communication sequence, and infer whether that stream is in an active or inactive dynamic regime. They do not model a multivariate marked network in which one shared latent episode simultaneously activates cross-excitation across a coordinated subset of actors and marks. In our setting, marks are part of the signal of coordination: they identify the channels through which the episode propagates, which makes the recovered structure more interpretable and more actionable in downstream analysis.

Change-point Hawkes methods, Koley et al. (2023) and Zhang et al. (2024), are also related but address different problems. Two technical distinctions deserve emphasis. First, change-point Hawkes estimate a partition of the timeline together with segment-specific parameters: each detected segment is fit independently, so evidence from one episode cannot be pooled with evidence from another. Our model instead carries a single active-regime parameter block, including sender/receiver gates, mark-channel weights, and the active-regime decay rate, shared across all active episodes, with the latent CTMC controlling only when that block is engaged. This sharing is what allows the conditional finite-sample guarantee in Theorem 3 to scale with cumulative active exposure $\pi_1 T$ rather than with the duration of any single episode, and it is precisely the regime in which short, recurrent bursts become statistically informative. Second, change-point methods produce hard boundaries: events immediately before and after a detected change-point belong to disjoint generative models. Our birth-time gating $Z(s)$ instead allows an active-born event to continue exciting downstream activity after the regime has switched off, which is the correct generative semantics when the structural signature of an episode resolves only through its consequences. Change-point Hawkes is therefore appropriate when the goal is to partition time into homogeneous segments; our model targets the complementary problem of recovering a recurring coordination structure that is intermittently activated. Neural temporal point-process models such as the neural Hawkes process (Mei & Eisner, 2017) and the Transformer Hawkes Process (Zuo et al., 2020) are highly expressive for nonlinear history dependence and next-event prediction. However, they typically do not impose a shared latent coordination field or a sparse, queryable directed interaction structure, so we view them as complementary predictive models rather than directly comparable approaches for interpretable episodic subgroup recovery.

Our problem also differs from classical link prediction in network analysis. Link prediction typically assumes a partially observed graph and assigns scores or probabilities to missing or future edges using observed topology, node attributes, or supervised classification. In contrast, no active coordination graph is observed in our setting. The directed edge set is an unobserved parameter of a marked temporal point process and must be inferred from event times, marks, and the compensator structure of the likelihood. Our approach is therefore closer to likelihood-based network topology inference: candidate edges are evaluated by whether they improve the generative explanation of the observed event stream. The additional challenge is that the relevant topology is episodic rather than persistent, so the model must infer both the latent activation field and the directed network activated by that field.

The distinction is therefore not that prior work lacks latent states, marks, or Hawkes excitation. It is that episodic directed coordination requires a recovery guarantee for an active network that is only intermittently expressed. Our contribution is to characterize when that network is identifiable, show how common modeling choices violate the required separation, and evaluate an estimator designed around the resulting recovery conditions.

3 Switching Marked Point Processes and Failure Modes

We study a class of switching marked point processes in which a latent field intermittently activates additional cross-excitation. This class is expressive enough to represent episodic coordination, but it also contains non-identifiable subfamilies. We note that if routine and active kernels have the same temporal shape, only their sum is identifiable. If mark-transition weights are unconstrained then actor couplings and mark

sensitivities can rescale each other without changing the likelihood. And if we gate excitation by the present regime rather than the birth-time regime of the parent event, downstream consequences of active-born events can be incorrectly attributed to the dormant state. We therefore introduce the model below together with the failure modes it is designed to avoid.

We observe K actors over a time horizon $[0, T]$. Actor i produces a marked event stream $\mathcal{D}_i = \{(t_n^{(i)}, m_n^{(i)})\}_{n=1}^{N_i}$, where $t_n^{(i)} \in (0, T]$ is the event birth time and $m_n^{(i)} \in \mathcal{M}$ is the event mark (e.g., a login or file access on a network, a spike from a recorded neuron, a transaction on an asset). We pool all events into a single time-ordered sequence $\mathcal{D} = \{(t_n, c_n, m_n)\}_{n=1}^{N}$, where $c_n \in \{1, \ldots, K\}$ denotes the actor and $m_n \in \mathcal{M}$ the mark. The model has two layers- a latent two-state process $Z(t)$ that switches between dormant and active coupling regimes, and a marked multivariate Hawkes process that governs the observed events. Routine excitation is always present, whereas the active regime adds an extra structured excitation channel. In inference, the observation window is discretized into intervals and each interval is scored under competing dormant and active explanations through the usual Hawkes log-intensity reward and compensator penalty.

Marked regime-switching Hawkes process. A hidden binary process $Z(t) \in \{0, 1\}$ governs whether cross-actor coupling is dormant ($Z = 0$) or active ($Z = 1$). It follows a two-state continuous-time Markov chain with activation rate η_{on} and deactivation rate η_{off}. The stationary active probability is $\pi_1 = \eta_{\text{on}}/(\eta_{\text{on}} + \eta_{\text{off}})$. We encode the prior for rare and episodic coordination through $\eta_{\text{on}} \ll \eta_{\text{off}}$, so $\pi_1 \ll 1$. Conditional on this latent field, the mark-specific intensity for actor i and mark m at current time t is

$$\lambda_i(t, m \mid \mathcal{H}_t, Z) = \nu_i(m) + \sum_{j=1}^{K} \sum_{m' \in \mathcal{M}} \int_0^{t^-} \phi_{j \to i}(t - s,\, m', m,\, Z(s))\, dN_j(s, m'), \tag{1}$$

where $\nu_i(m) \geq 0$ is the mark-specific baseline intensity, t is the current time, s denotes the time of a past event, and $dN_j(s, m')$ places mass at past mark-m' events of actor j. The kernel $\phi_{j \to i}(t - s, m', m, Z(s))$ is the contribution of a past event from actor j with mark m' at time s to the current intensity of actor i and mark m at time t. Its arguments encode the time lag $t - s$, the parent mark m', the child mark m, and the latent regime $Z(s)$ at the time the parent event occurred. Thus the integral sums the contributions of all past such events to the current intensity.

Mark-to-mark channels preserve structural information that would be lost in an unmarked event stream. In an unmarked model, the active regime can only appear through changes in actor-to-actor excitation or event rate. This makes episodic activity look like a stronger or weaker version of the same kernel that already captures routine bursting. With marks, the active regime can instead be characterized by changes in event-type transitions. For example, a routine sequence may involve a login followed by ordinary file access, while an active episode may involve privilege escalation followed by a burst of downloads. In financial data, buyer- and seller-initiated trades can alternate differently during crisis periods than during routine market activity. In neural data, different event types can similarly encode stimulus- or state-dependent transitions. These mark pairs carry directional information that is not reducible to event timing alone. We therefore use mark-to-mark channels to preserve the structural signature of episodic activity. In §6.7, we empirically test this design choice on a financial spike-in and find that the active regime reshapes the recovered buy/sell transition matrix rather than merely increasing event volume. The benefit of marks is still data-dependent, but the FTX spike-in experiments show that marks can carry recoverable coordination evidence distinct from actor membership and aggregate rate.

Birth-time vs. Present-time gating. Since the kernel depends on $Z(s)$, the regime at the time the past event occurred, it allows an event born during an active episode to carry its regime label forward and continue to excite other actors even after the field switches off. A burst of collusive coordination can therefore generate a cascade that outlasts the episode itself. The kernel decomposes into routine and active components:

$$\phi_{j \to i}(\tau, m', m, Z(s)) = \alpha_{j \to i}^{(0)}\, \rho_{m'm}^{(0)}\, \kappa_0(\tau) + Z(s)\, \alpha_{j \to i}^{(1)}\, \rho_{m'm}^{(1)}\, \kappa_1(\tau), \tag{2}$$

where $\tau = t - s$ is the lag between the parent event and the current time, and $\int_0^\infty \kappa_r(\tau)\, d\tau = 1$. The first term captures always-on routine coupling, such as shared systems, team workflows, or reporting lines. The second

activates only when the originating event was stamped $Z(s) = 1$. The mark-transition weights $\rho^{(r)}_{m'm} \geq 0$ capture mark-to-mark triggering under regime $r \in \{0, 1\}$, for example, a privilege-escalation event may trigger downstream file-download events more strongly than routine access events. We use exponential temporal kernels

$$\kappa_r(\tau) = \beta_r e^{-\beta_r \tau} \mathbf{1}\{\tau \geq 0\}, \qquad \beta_r > 0, \quad r \in \{0, 1\},$$

where β_0 and β_1 are the dormant- and active-regime decay rates, respectively. This choice preserves the normalization $\int_0^\infty \kappa_r(\tau)\, d\tau = 1$ and permits recursive trace updates and $O(N)$ likelihood evaluation.

Actor-level gates and directed interaction scores. The model separates two distinct mechanisms: an actor's overall membership in a coordinated burst, and their specific pairwise interactions with other members. We encode membership using nonnegative sender and receiver gates ($g_j, h_i \geq 0$), which act as strict filters for participation. Conditional on passing these gates, the directed interaction score is then governed by low-rank embeddings $\mathbf{u}_j, \mathbf{v}_i \in \mathbb{R}^d$ with $d \ll K$, which control directed interaction strength within the active subgroup. The active coupling is parameterized as

$$\alpha^{(1)}_{j\to i} = g_j\, h_i\, \mathrm{softplus}(\mathbf{u}_j^\top \mathbf{v}_i). \tag{3}$$

When $g_j = 0$, actor j cannot send active excitation regardless of the embeddings, when $h_i = 0$, actor i cannot receive it. The softplus ensures positive within-subgroup coupling, and the asymmetry $\mathbf{u}_j \neq \mathbf{v}_j$ encodes directionality, so a hub can have a large sender embedding and a small receiver embedding. The directed pair score $s_{j\to i} = g_j h_i\, \mathrm{softplus}(\mathbf{u}_j^\top \mathbf{v}_i)$ is invariant to the sender–receiver gate rescaling and to embedding-coordinate transformations that preserve the inner products $\mathbf{u}_j^\top \mathbf{v}_i$. Accordingly, edge-level evaluation uses $s_{j\to i}$ rather than the raw embedding coordinates. The embeddings themselves are identifiable only up to these transformations; the recovery results in section §6 concern gate supports and active couplings, not the coordinates of $\mathbf{u}_j$ and $\mathbf{v}_i$ individually. For the routine coupling $\alpha^{(0)}_{j\to i}$, we use a scalar parameter per actor pair, regularized toward zero. In applications with known team structure, these terms can be constrained by reporting or adjacency relations, otherwise they are treated as nuisance parameters with ℓ_2 regularization, serving strictly to absorb routine background variance so it does not confound the active structure.

We construct a penalized objective to balance data fit against structural parsimony. Let Θ collect all model parameters: $\Theta = \big(\{\nu_i(m)\},\ \{\alpha^{(0)}_{j\to i}\},\ \{g_j, h_i, \mathbf{u}_j, \mathbf{v}_i\},\ \{\rho^{(r)}_{m'm}\},\ \beta_0, \beta_1,\ \eta_{\mathrm{on}}, \eta_{\mathrm{off}}\big)$. We estimate Θ by maximizing a penalized marginal log-likelihood: $\widehat{\Theta} = \arg\max_\Theta \big\{\log p(\mathcal{D} \mid \Theta) - \Omega(\Theta)\big\}$, with regularizer

$$\Omega(\Theta) = \lambda_g \sum_{j=1}^K g_j + \lambda_h \sum_{i=1}^K h_i + \lambda_e \Big(\sum_{j=1}^K \|\mathbf{u}_j\|_2 + \sum_{i=1}^K \|\mathbf{v}_i\|_2\Big) + \lambda_0 \sum_{j,i} (\alpha^{(0)}_{j\to i})^2. \tag{4}$$

The first two terms impose ℓ_1 sparsity on the scalar gates, producing a hard partition between subgroup members and non-members. The third applies a group-ℓ_2 penalty to the embeddings, stabilizing the low-rank structure among surviving actors. The fourth applies ℓ_2 shrinkage to the routine coupling block, which is treated as a nuisance component.

Penalty-balanced gate gauge. The sender–receiver gate factorization in Eq. 3 has a one-dimensional scaling gauge: for any $c > 0$, the transformation $g_j \mapsto c g_j$ and $h_i \mapsto c^{-1} h_i$ leaves every active coupling $\alpha^{(1)}_{j\to i}$, and hence the likelihood and the gate supports, unchanged. Because the gate penalty in Eq. 4 is not invariant to this transformation, the penalized objective selects the representative that minimizes the gate penalty along the orbit. Assuming $\lambda_g, \lambda_h > 0$ and nonempty sender and receiver supports, we define the penalty-balanced representative $(g^{\mathrm{pb}}, h^{\mathrm{pb}})$ by

$$g_j^{\mathrm{pb}} = c_{\mathrm{pb}} g_j, \qquad h_i^{\mathrm{pb}} = c_{\mathrm{pb}}^{-1} h_i, \qquad c_{\mathrm{pb}} = \left(\frac{\lambda_h \sum_i h_i}{\lambda_g \sum_j g_j}\right)^{1/2}.$$

Equivalently, this is the unique positive representative satisfying

$$\lambda_g \sum_{j=1}^K g_j^{\mathrm{pb}} = \lambda_h \sum_{i=1}^K h_i^{\mathrm{pb}}.$$

Indeed, along the orbit the gate penalty is $\lambda_g c \sum_j g_j + \lambda_h c^{-1} \sum_i h_i$, whose unique minimizer is c_{pb}. Thus any global maximizer of the penalized objective, and the reported output of the estimator after the harmless orbit-rescaling above, lives in the same penalty-balanced gauge. All gate-level beta-min quantities in Theorem 3 are measured in this gauge.

4 Identifiability and Recovery

We describe next the population and finite-sample conditions under which episodic coordination is identifiable and recoverable. The results separate three questions that are conflated in standard switching Hawkes formulations: whether the active marked kernel can be separated from routine excitation, whether actor-level couplings can be separated from mark-transition weights, and whether a posterior-weighted estimator recovers the true sender and receiver supports from finite active exposure. The latent-state results below should be read conditionally: Theorem 1 assumes observed birth-state labels or a posterior that is consistent for those labels, and Theorem 3 assumes the implemented E-step surrogate remains uniformly close to the oracle birth-state posterior. These posterior-accuracy requirements are not derived from primitive mixing and emission conditions in this paper; they are explicit assumptions that separate the statistical recovery argument from the latent-state approximation problem.

Assumption 1. *Stability Condition. Define* $\mathbf{A}^{(r)}_{ji} = \alpha^{(r)}_{j\to i} \sum_{m',m} \rho^{(r)}_{m'm}$. *For the process to be non-explosive, the integrated kernel matrix in the active regime must have spectral radius less than one. When the active regime is on, total branching is governed by* $\mathbf{A}^{(0)} + \mathbf{A}^{(1)}$, *since dormant excitation remains present. For non-explosivity, this combined matrix must have spectral radius less than one. Stability in the dormant regime follows automatically.*

Theorem 1 (Active-kernel identifiability under birth-state separability)**.** *Assume the stability condition in Assumption 1. Consider the switching marked Hawkes model in Eq. 1–2 under either observed birth-state labels or a latent regime posterior that is consistent for the birth-state labels in the sense that the posterior error vanishes uniformly over event-containing intervals. Then the active marked kernel block*

$$\phi^{(1)}_{(j,m')\to(i,m)}(\tau) := \alpha^{(1)}_{j\to i}\rho^{(1)}_{m'm}\kappa_1(\tau)$$

is identifiable from the law of the marked process provided:

(a) both latent states are visited with positive probability, $\eta_{\mathrm{on}} > 0$ *and* $\eta_{\mathrm{off}} > 0$*;*
(b) active episodes generate nondegenerate exposure from coupled actors;
(c) dormant–active separation is available either through observed birth-state labels, through consistently recovered birth-state labels (which is itself the posterior-accuracy/exposure requirement formalized in Assumption 2), or, when birth-state information is not available, through linearly independent routine and active temporal kernels, $\kappa_0 \not\propto \kappa_1$*, which for exponential kernels is equivalent to* $\beta_0 \neq \beta_1$.

The active actor-level couplings $\alpha^{(1)}_{j\to i}$ *are identifiable from this marked kernel block if the active mark-transition weights* $\rho^{(1)}_{m'm}$ *are known or are identified under a normalization such as*

$$\sum_{m\in\mathcal{M}} \rho^{(1)}_{m'm} = 1 \qquad \text{for each } m' \in \mathcal{M}.$$

See Appendix A.1 for a Proof. Conditions (a) and (b) are exposure requirements: the process must enter the active state and produce enough active-born events for the active block to leave a statistical trace. Condition (c) states the separation route. If birth-state labels are observed or consistently recovered, dormant-born parents identify the dormant excitation level and active-born parents identify the active excitation level, so the active increment is identified by their difference. For latent labels, however, "consistently recovered" is not a free alternative: under collinearity it is the only remaining route to separation and carries the posterior-accuracy and exposure requirements made explicit in Assumption 2 and Theorem 3. If birth-state information is not available, temporal-shape separation can provide an alternative route; when the model is collapsed to a birth-state-blind single-kernel representation and the kernels are collinear, Theorem 2 shows the routine/active split is not identifiable from that collapsed summary. The final normalization requirement

is an actor–mark-separation requirement: without an anchor on $\rho^{(1)}$, actor coupling and mark sensitivity are identifiable only up to reciprocal scaling. Figure 1 illustrates the practical implications of Theorem 1: any point on the unnormalized contour produces an identical observed cascade, so the true network is impossible to resolve without an anchor. We fix the mark weights to force the actor couplings to their single true value, guaranteeing structural identifiability.

Assume the temporal kernels are normalized, $\int_0^\infty \kappa_r(\tau)\, d\tau = 1$. If the routine and active temporal kernels are collinear, then $\kappa_0 = \kappa_1 =: \kappa$. In this case the birth-time-gated model has two excitation levels,

$$\lambda_i(t, m \mid \mathcal{H}_t, Z) = \nu_i(m) + \sum_{j,m'} \int_0^{t^-} \Big[B^{(0)}_{(j,m')\to(i,m)} + Z(s) B^{(1)}_{(j,m')\to(i,m)} \Big] \kappa(t-s)\, dN_j(s, m'),$$

so a dormant-born parent carries excitation level $B^{(0)}$, while an active-born parent carries excitation level $B^{(0)} + B^{(1)}$.

Theorem 2 (Collinearity and birth-state-blind collapse)**.** *Fix any $q \in (0, 1]$ and consider the restricted birth-state-blind, time-invariant single-kernel representation whose population kernel-level object is*

$$B^{\text{mix}}_q = B^{(0)} + qB^{(1)}.$$

Within this restricted collapsed representation, the active block $B^{(1)}$ is not identifiable from B^{mix}_q, even inside the structured Kronecker family $B^{(r)} = A^{(r)} \otimes R^{(r)}$. In particular, there exist two distinct stable structured parameterizations with the same B^{mix}_q and different active blocks. Consequently, no estimator whose input is restricted to this single-kernel collapsed representation can recover $B^{(1)}$ uniformly over the structured class.

See Appendix A.2 for a proof.

Remark 1 (Scope of the collinearity result)**.** *Theorem 2 is intentionally scoped to a birth-state-blind collapsed representation. It does not claim that every collinear switching Hawkes model is non-identifiable from the full observed law. With birth-time gating, the common-kernel model still has two excitation levels: a dormant-born parent carries $B^{(0)}$, while an active-born parent carries $B^{(0)} + B^{(1)}$. Thus, if birth-state labels are observed, or if the birth-state posterior is consistent and both states have nondegenerate exposure, the active block is identifiable by subtraction, $B^{(1)} = (B^{(0)} + B^{(1)}) - B^{(0)}$. What collinearity removes is the independent temporal-shape contrast between routine and active excitation. If birth states are not recovered, a static or birth-state-blind estimator sees only the collapsed object B^{mix}_q, and the routine/active decomposition is not identifiable from that summary. This is why Theorem 3 explicitly assumes posterior accuracy, and why the shared-decay/mark-channel ablation ($\beta_0 = \beta_1,\ \rho^{(0)} = \rho^{(1)}$) is interpreted as a collapse of temporal and mark-channel separation rather than as an unconditional infinite-sample impossibility.*

Proposition 1 (Actor–mark scaling ambiguity)**.** *Let $B^{(1)}_{(j,m')\to(i,m)} = \alpha^{(1)}_{j\to i}\rho^{(1)}_{m'm}$ denote the active marked excitation block. For any $c > 0$, the parameter pair*

$$\tilde{\alpha}^{(1)}_{j\to i} = c\alpha^{(1)}_{j\to i}, \qquad \tilde{\rho}^{(1)}_{m'm} = c^{-1}\rho^{(1)}_{m'm}$$

induces the same $B^{(1)}$. Hence actor couplings and mark-transition weights are not individually identifiable without a normalization or external anchor on $\rho^{(1)}$.

Support recovery under posterior-accurate EM. We now provide a conditional result linking active exposure time to subgroup recovery. Let $\bar{\mu}$ be a lower bound on the aggregate conditional event intensity contributed by true subgroup actors during active intervals (so that larger $\bar{\mu}$ means more active-born evidence accrues per unit of active exposure from the subgroup as a whole). Equivalently, if one works with a per-actor lower bound $\bar{\mu}_{\text{act}}$, then the aggregate convention corresponds to $\bar{\mu} \asymp k\bar{\mu}_{\text{act}}$ up to constants. Let $S^\star_g := \{j : g^\star_j > 0\}$, $S^\star_h := \{i : h^\star_i > 0\}$, and $S^\star := S^\star_g \cup S^\star_h$ denote the true sets of senders, receivers, and total participants (the target subgroup), with size $k := |S^\star|$ and effective complexity $s_\theta \asymp kd$, where d is per-edge parameter dimension. Define the subgroup's quietest active coupling and penalty-gauge gate margins as

$$\alpha_{\min} := \min_{j\in S^\star_g,\ i\in S^\star_h} \alpha^{(1),\star}_{j\to i}, \qquad g^{\text{pb}}_{\min} := \min_{j\in S^\star_g} g^{\text{pb},\star}_j, \qquad h^{\text{pb}}_{\min} := \min_{i\in S^\star_h} h^{\text{pb},\star}_i,$$

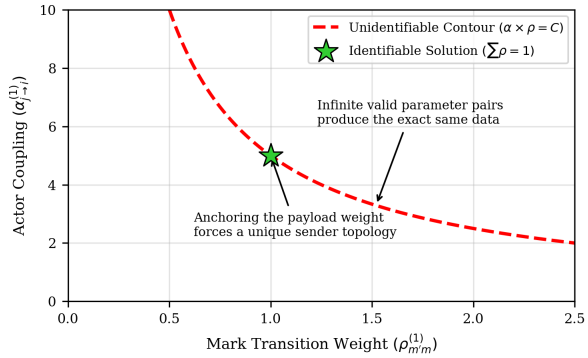

Figure 1: Illustration of the scaling ambiguity addressed in Theorem 1. Without normalization, the product of actor coupling and mark sensitivity forms a contour of infinite equivalent solutions (dashed red). Normalizing the mark weights gives structural identifiability under the stated normalization.

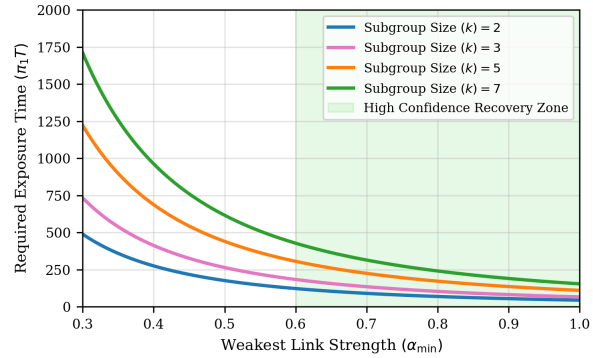

Figure 2: Required active exposure time versus weakest link strength at fixed aggregate subgroup active-event rate, illustrating Theorem 3. Weak participants require quadratically more active exposure; larger active parameter blocks shift the requirement through s_θ.

where $(g^{\mathrm{pb},\star}, h^{\mathrm{pb},\star})$ is the penalty-balanced representative defined after equation 4. Finally, define the effective active sample size

$$N_{\mathrm{eff}} := \sum_{n=1}^{N} \mathbf{1}[Z(t_n) = 1]\, \mathbf{1}[c_n \in S^\star],$$

the number of active-born events produced by true subgroup actors.

Assumption 2 (Recovery conditions, including posterior accuracy)**.** *Suppose the following hold:*

1. *the stability and identifiability conditions of Theorem 1;*
2. *the oracle active-block surrogate satisfies a local restricted-curvature condition on the active parameter block and a coordinatewise score bound of order*

$$\lambda_T = C_0 \sqrt{\frac{\log(K^2/\delta)}{N_{\mathrm{eff}}}};$$

3. *the active exposure and effective active event count concentrate at their nominal rates, so that $\pi_1 T$ active time yields N_{eff} of order $\bar{\mu}\,\pi_1 T$, as made explicit in Appendix A.3;*
4. *after initialization, the posterior-weighted E-step is uniformly close to the oracle birth-state posterior, with approximation error $o(\lambda_T)$, so that the implemented surrogate is a vanishing perturbation of the oracle surrogate.*

Theorem 3 (Conditional finite-sample support recovery under posterior accuracy)**.** *Given Assumption 2, there exist constants $C_1, C_2 > 0$ such that if*

$$\pi_1 T \geq C_1 \frac{s_\theta \log(K^2/\delta)}{\alpha_{\min}^2 \bar{\mu}}, \tag{5}$$

and $g^{\mathrm{pb}}_{\min} \wedge h^{\mathrm{pb}}_{\min} \geq C_2 \lambda_T$, then, with probability at least $1-\delta$, $\widehat{S}_g = S^\star_g, \widehat{S}_h = S^\star_h$. Consequently, the recovered participant set $\widehat{S} := \widehat{S}_g \cup \widehat{S}_h$ equals $S^\star$.

See Appendix A.3 for a proof. This result is conditional on posterior accuracy and does not claim global convergence of EM from arbitrary initialization, nor does it prove that the posterior-weighted mean-field-style surrogate achieves this accuracy from primitive emission and mixing assumptions. We empirically probe this posterior-accuracy bridge in the controlled synthetic setting in §6, but do not claim to prove posterior concentration from primitive conditions.

Table 1: Identifiability map for episodic coordination in switching marked point processes. The estimator is designed to repair these failure modes rather than merely improve prediction; the finite-sample support-recovery theorem is conditional on posterior accuracy of the E-step.

Failure mode	**Statistical problem**	**Repair**	**Evidence in paper**
No active exposure	The active state is visited too rarely, or active-born events from the true participants are too few.	Effective exposure condition based on N_{eff} (birth-time events) and $\pi_1 T$.	Theorem 3; exposure experiments.
Kernel collinearity with birth-state-blind fitting	A collapsed single-kernel model cannot split routine from active excitation when temporal shapes coincide.	Use recoverable birth states or non-collinear kernels, with $\beta_0 \neq \beta_1$ for exponentials.	Theorem 2; D4 collapse.
Posterior inaccuracy	The active support proof fails if the E-step posterior is not close to the oracle birth-state posterior.	Assume and diagnose rather than design-repair: inspect activation-posterior bands and, where oracle labels exist, exposure/collapse diagnostics. Raw CA1/FTX claims remain conditional on Assumption 2, item 4 (Condition C6 in Appendix A.3).	Assumption 2; Theorem 3; Fig. 5.
Actor–mark scaling ambiguity	$\alpha^{(1)}_{j\to i}\rho^{(1)}_{m'm}$ is invariant to reciprocal rescaling.	Normalize or fix mark-transition weights.	Proposition 1; Fig. 1.
Birth-time gating	Consequences of active-born events are reassigned when the regime switches off.	Gate excitation by parent event birth time $Z(s)$.	D1 ablation (§6.3); hub recovery collapse.
Segmentation without assignment	An HMM can identify active windows but assign the active actors or roles incorrectly.	Joint inference of latent field and active network.	B2 synthetic (Table 3) and CA1 spike-in results.

The theorem's model-specific content is the identification of N_{eff} as the right exposure scale for the birth-time-gated support problem; the remaining proof machinery follows standard high-dimensional point-process and M-estimation arguments under the oracle regularity conditions. We separate the statistical recovery problem from the latent-state approximation problem: if the E-step assigns birth-state mass close enough to the oracle posterior, then the active support is recovered at the exposure rate in Eq. 5. We test this regime using exposure and collapse experiments in the results section. Figure 2 illustrates the practical bound of Theorem 3. The required observation window is bottlenecked by the faintness of the subgroup: as the weakest-link strength ($\alpha_{\min}$) becomes quieter, the necessary exposure time grows quadratically. The background network size K enters through the logarithmic multiplicity term, while subgroup size enters through $s_\theta \asymp kd$ when the aggregate subgroup active-event rate $\bar{\mu}$ is held fixed. If one instead parameterizes exposure by a per-actor rate, the corresponding aggregate rate scales as $k\bar{\mu}_{\text{act}}$ and the leading k factors cancel; throughout the theorem we use the aggregate-rate convention for the active-duration statement.

5 Estimator

The theory above identifies the quantities that must be preserved for recovery: birth-state attribution, dormant–active separation, normalized mark-transition structure, and sparse sender and receiver support (Table 1). The estimator is built to approximate these quantities when the latent regime path is unobserved. Direct maximization of the penalized marginal objective is intractable because birth-time gating causes past regime labels to propagate forward through the intensity. We therefore use a generalized EM procedure that alternates between posterior-weighted inference of the activation field and regularized updates of the active

network. This is an approximate surrogate algorithm, not an exact optimizer of the marginal likelihood; the finite-sample support guarantee in Theorem 3 applies only when its E-step posterior is close enough to the oracle birth-state posterior.

We partition the observation window into B intervals $I_b = (\tau_{b-1}, \tau_b], \qquad b = 1, \ldots, B$, with widths $\Delta_b = \tau_b - \tau_{b-1}$, and approximate the latent CTMC by a piecewise-constant state $Z_b \in \{0, 1\}$ over each interval. The resulting complete-data objective is

$$\log p(\mathcal{D}, Z \mid \Theta) = \log p(Z \mid \eta_{\text{on}}, \eta_{\text{off}}) + \log p(\mathcal{D} \mid Z, \Theta). \tag{6}$$

A full latent-parent derivation and the detailed trace construction are given in Appendix B.

Estimator as identifiability repair. The posterior-weighted trace construction preserves the distinction required by Theorem 1. Current events in interval I_b are evaluated under the current regime hypothesis z, while past events retain the posterior birth-state mass assigned to the interval in which they occurred. Thus an event that was likely born during an active episode continues to contribute active excitation downstream even if the current interval is evaluated as dormant. This is the computational analogue of birth-time gating in Eq. 2: the regime label is attached to the parent event, not overwritten by the regime at the child event's time. Without this bookkeeping, delayed consequences of an active-born event would be reassigned to the present regime, producing exactly the present-time attribution failure tested by the D1 design ablation (§6.3).

E-step. Given that the kernel is gated at the *originating event birth time*, the emission likelihood for interval I_b depends on earlier latent states as well as the current one. Exact inference over all regime paths is therefore exponential in B. To achieve tractability, we use a posterior-weighted approximation in which pre-interval active contributions are replaced by their previous-iterate posterior weights, yielding a trace-based approximation to the marked intensity. Therefore, the E-step needs two history summaries for each actor–mark pair (j, m'): one for routine excitation and one for active excitation. The routine trace collects all past events of mark m' by actor j, decayed under the routine kernel:

$$B_{j,m'}(t) = \sum_{\ell:\, t_\ell < t,\ c_\ell = j,\ m_\ell = m'} \kappa_0(t - t_\ell).$$

This trace is always present, because routine excitation is not gated by the latent regime.

The active trace has the same form, but each past event is weighted by the probability that it was born in the active regime. When evaluating the current interval I_b under regime hypothesis $Z_b = z$, we define

$$\widetilde{A}^{(t-1)}_{j,m'}(t, z) = \sum_{\ell:\, t_\ell < t,\ c_\ell = j,\ m_\ell = m'} \widetilde{w}^{(t-1)}_\ell(t, z)\, \kappa_1(t - t_\ell),$$

where

$$\widetilde{w}^{(t-1)}_\ell(t, z) = \begin{cases} z, & t_\ell \in I_b, \\ \gamma^{(t-1)}_{b_\ell}(1), & t_\ell \notin I_b. \end{cases}$$

Thus events born in the current interval inherit the regime hypothesis being tested, while earlier events carry the active-state posterior previously assigned to their own birth interval. This is the posterior-weighted implementation of birth-time gating: *active excitation follows the parent event forward in time rather than being reassigned according to the regime active at the child event's time.* Both traces admit $O(1)$ recursive updates per event under exponential kernels. The approximate intensity then is

$$\widetilde{\lambda}_i(t, m \mid Z_b = z) = \nu_i(m) + \sum_{j,m'} \alpha^{(0)}_{j\to i} \rho^{(0)}_{m'm} B_{j,m'}(t) + \sum_{j,m'} \alpha^{(1)}_{j\to i} \rho^{(1)}_{m'm} \widetilde{A}^{(t-1)}_{j,m'}(t, z). \tag{7}$$

This yields the interval emission log-likelihood combining terms for the marked event intensity $\widetilde{\lambda}_{c_n}$ (see Eq. 24) and $\widetilde{\lambda}_i$, balancing the occurrence of events against the intervening periods of inactivity.

$$\ell_b(z) = \sum_{n:\, t_n \in I_b} \log \widetilde{\lambda}_{c_n}(t_n, m_n \mid Z_b = z) - \sum_{i=1}^{K} \sum_{m \in \mathcal{M}} \int_{I_b} \widetilde{\lambda}_i(s, m \mid Z_b = z)\, ds. \tag{8}$$

The first term accumulates the log-intensity exactly at the observed event birth times (rewarding accurate temporal prediction), while the integral term acts as the log-survival probability across all inter-event gaps. If the model overestimates the marked intensity during periods of actual silence, the survival probability collapses, heavily penalizing the surrogate objective. Using these emissions together with the CTMC transition matrix induced by $(\eta_{\text{on}}, \eta_{\text{off}}, \Delta_b)$, we run a standard forward-backward recursion to obtain

$$\gamma_b^{(t)}(z) = p(Z_b = z \mid \mathcal{D}, \Theta^{(t)}), \tag{9}$$

$$\xi_b^{(t)}(z, z') = p(Z_b = z, Z_{b+1} = z' \mid \mathcal{D}, \Theta^{(t)}). \tag{10}$$

The sequence $\gamma_b^{(t)}(1)$ is the inferred activation timeline.

M-step. Given (γ, ξ), we increase the generalized EM surrogate

$$Q(\Theta \mid \Theta^{(t)}) = \sum_{b=1}^{B} \sum_{z \in \{0,1\}} \gamma_b^{(t)}(z)\, \ell_b(z) + \sum_{b=1}^{B-1} \sum_{z,z'} \xi_b^{(t)}(z, z') \log P_{zz'} - \Omega(\Theta). \tag{11}$$

The CTMC rates are updated in closed form from expected sojourn times and transition counts, followed by clipping to a numerically stable range. The mark-specific baseline rates are updated in closed form: for each actor i and mark m, $\nu_i(m)$ is set to the expected number of events of type (i, m) attributed to the baseline component, divided by the observation horizon T. The active block (g, h, U, V) is then updated by proximal gradient ascent on a trace-based surrogate: the sender and receiver gates use nonnegative ℓ_1 proximal steps, while the sender and receiver embedding rows use group-ℓ_2 shrinkage after a gradient step. The remaining nuisance parameters $(A^{(0)}, \rho^{(0)}, \rho^{(1)}, \beta_0, \beta_1)$ are updated by projected generalized-EM ascent with backtracking on the recomputed surrogate. For a detailed derivation of the inference procedure see Appendix B.

Initialization and output. We initialize the routine and active excitation parameters with small feasible values, choose admissible decay and CTMC transition rates, and initialize the active-state posterior to a weakly active prior before refining it by generalized EM. The fitted model returns the activation timeline $\gamma_b(1)$, the sender and receiver gate supports, and the directional within-subgroup scores $\widehat{s}_{j \to i} = \widehat{g}_j \widehat{h}_i \,\text{softplus}(\widehat{\mathbf{u}}_j^\top \widehat{\mathbf{v}}_i)$.

Algorithm 1 shows the inference procedure.

Algorithm 1 Regime-switching Hawkes EM

1: Initialize $\Theta^{(0)}$ from empirical event statistics and small random parameter values, initialize $\gamma^{(0)}$ to a weakly active prior.
2: **for** $t = 1, 2, \ldots$ until convergence **do**
3: **E-step:** build posterior-weighted routine and active traces, evaluate interval emissions $\ell_b(z)$, and run forward-backward to obtain $\gamma^{(t)}$ and $\xi^{(t)}$.
4: **M-step:** update CTMC rates in closed form, refresh baseline intensities in closed form, update the active block (g, h, U, V) by proximal generalized-EM ascent, update $(A^{(0)}, \rho^{(0)}, \rho^{(1)}, \beta_0, \beta_1)$ by projected generalized-EM ascent.
5: Apply penalty annealing, momentum blending, and stability control.
6: **end for**
7: **Output:** activation timeline, inferred subgroup members, and directional interaction scores.

6 Empirical Recovery and Identifiability Stress Tests

Theorem 3 makes four empirical predictions in observable coordinates, so we organize the experiments by prediction rather than by dataset. (P1) asks whether recovery improves with realized effective active exposure,

$$N_{\text{eff}} = \sum_{n=1}^{N} \mathbf{1}[Z(t_n) = 1] \mathbf{1}[c_n \in S^\star],$$

Table 2: Theory-facing empirical design.

Testbed	**Experiment (P1) (Exposure)**	**Experiment (P2) (Weak Link)**	**Experiment (P3) (Posterior)**	**Experiment (P4) (Collapse)**
Synthetic	**Evaluate** known-truth recovery versus active-born event count; partition partial and perfect seeds by N_{eff}.	**Sweep** weakest-link parameter in a controlled setting; extract the fitted log–log threshold slope.	**Execute** oracle-posterior and posterior-corruption stress tests to measure R_γ.	**Ablate** (D1–D4) birth-time gating, role separation, low-rank structure, and temporal/mark separation.
UCDP GED	**Inject** volume-neutral real-background spike-ins; vary realized planted N_{eff} against fixed background.	**Modulate** the planted share of one designated weak receiver while holding aggregate active rate fixed.	**Calculate** event-weighted gap from $\widehat{\gamma}_b(1)$ to planted active windows; report by R_γ band.	**Contrast** with B1 static-gate to rule out actor-volume selection; apply B2 for segmentation-without-assignment.
CRCNS CA1	**Re-index** existing spike-ins by realized active-born events from planted units.	**Sweep** weak-unit inclusion thresholds (optional).	**Measure** R_γ against planted active windows in spike-ins; raw CA1 remains conditional.	**Compare** raw CA1 decay-scale separation and synthetic D4 against B1/B2/B3 collapse baselines.
FTX	**Track** dose-response by re-indexing receiver-target recovery and mark-channel lift against realized N_{eff}.	**Omit** from controlled sweep; a single targeted cascade does not isolate the weakest link.	**Report** R_γ against planted windows if target is retained.	**Assess** qualitative marked-channel representation compared to unmarked limitations.

Table 3: Synthetic subgroup recovery against internal ablations and an external change-point Hawkes baseline. Values report member-set F1, mean ± standard deviation over 20 seeds.

Method	Base benchmark $K = 10,\ \lvert S^\star\rvert = 3$	Scaled benchmark $K = 20,\ \lvert S^\star\rvert = 4$
Proposed	0.867 ± 0.274	0.975 ± 0.077
B1: Static Hawkes + gates	0.633 ± 0.284	0.750 ± 0.181
B2: Modular HMM + Hawkes	0.283 ± 0.271	0.175 ± 0.169
B3: Static Hawkes + SVD	0.333 ± 0.000	0.250 ± 0.000
B4: Change-point Hawkes	0.733 ± 0.256	0.975 ± 0.077

rather than with total event volume or nominal injection strength. (P2) asks whether, at fixed aggregate active-event rate, the active exposure time needed to recover a weak link scales as $1/\alpha_{\min}^2$. (P3) asks whether finite-sample recovery is conditional on the accuracy of the fitted active-state posterior at the λ_T scale. (P4) asks whether removing the model ingredients that create separability—birth-time attribution, role separation, temporal/mark separation, or joint activation-network inference—produces the predicted collapse modes. All spike-in experiments are reported in these theorem-facing coordinates. We record realized N_{eff}, weakest-link participation, active sojourn $\pi_1 T$, posterior gap R_γ, member-set F1, receiver F1, hub accuracy, and exact support recovery where available. Synthetic experiments provide the cleanest known-truth tests because the latent path, active block, and generating parameters are known.

Table 4 summarizes the empirical architecture: each experiment is assigned to the theorem-facing prediction it is meant to test, rather than treating every dataset as evidence for every condition.

Table 4: Prediction-by-dataset evidence matrix. Each row states one empirical prediction, the theorem condition it targets, and the evidence layer used to test it.

Prediction	Question	Theory anchor	Main evidence layer
(P1): Active exposure	Does recovery improve with realized active-born evidence rather than total event volume?	Theorem 3: recovery is controlled by the birth-time effective sample size N_{eff} and the active-duration condition.	Synthetic known-truth recovery, UCDP GED volume-neutral planted coordination, and CRCNS CA1 planted assembly; see §6.4.
(P2): Weakest link	At fixed aggregate active mass, do weaker links require longer active sojourns?	Theorem 3: the active-duration bound scales as $1/\alpha_{\min}^2$.	Synthetic true-$\alpha_{\min}$ weakest-link sweep with fixed aggregate receiver-directed active mass; see §6.5.
(P3): Posterior bridge	Does recovery depend on the fitted active-state posterior being close to the oracle birth-state posterior?	Assumption 2, item 4 (Condition C6 in Appendix A.3): the oracle-to-approximate transfer requires local E-step posterior accuracy.	Synthetic oracle-posterior comparison and posterior-corruption stress tests; see §6.6.
(P4): Separation / collapse	Do the predicted failures appear when an identifiability repair is removed?	Theorems 1, 2, and Proposition 1: recovery fails when active exposure, birth-state attribution, dormant–active separation, or actor–mark normalization is removed.	Synthetic ablations, D4 kernel-collapse diagnostic, CA1 segmentation-without-assignment, and FTX mark-channel recovery; see §6.3 and §6.7.

6.1 Datasets and real-background construction

The empirical section uses four kinds of testbeds: synthetic known-truth experiments, UCDP GED conflict-event backgrounds, CRCNS hippocampal CA1 recordings, and Binance trade-event data from the FTX-collapse window. The synthetic experiments are the theorem-clean setting: the latent regime path, active subgroup, sender/receiver roles, marks, and generating parameters are known. They are therefore used for the sharpest tests of exposure, weakest-link scaling, posterior accuracy, and ablation collapse. The real-background experiments are not used to make causal claims about conflict, neuroscience, or financial markets; they test whether the same identifiability logic survives on empirical event streams with realistic burstiness and heterogeneous background activity.

UCDP GED conflict-event background. The UCDP Georeferenced Event Dataset (GED) is an expert-coded event dataset of organized violence, with events disaggregated in time and geographic location and linked to UCDP conflict, dyad, country, and violence-type identifiers (Sundberg & Melander, 2013; Shawn Davies, 2026). We use GED as a real social-science event background because its records are naturally timestamped, actor-linked, and marked: each event includes a date, actor-side or dyad identifiers, location, type of violence, and fatality estimate. In our preprocessing, we restrict to state-based events with exact dates, select a balanced India window (2007–2011), map the top $K = 8$ dyad streams to point-process actors, and use fatality-severity bins derived from GED's best fatality estimate as marks: low (1), medium (2–4), and high (≥ 5). The resulting experiment is not a substantive model of conflict escalation; it is a real-background identifiability test in which planted episodic coordination must be recovered against an empirical social-science event stream. The main UCDP test is volume-neutral: we replace or reallocate events so that marginal actor volume is not sufficient to identify the planted subgroup, making it a direct test of birth-time active exposure rather than raw event count.

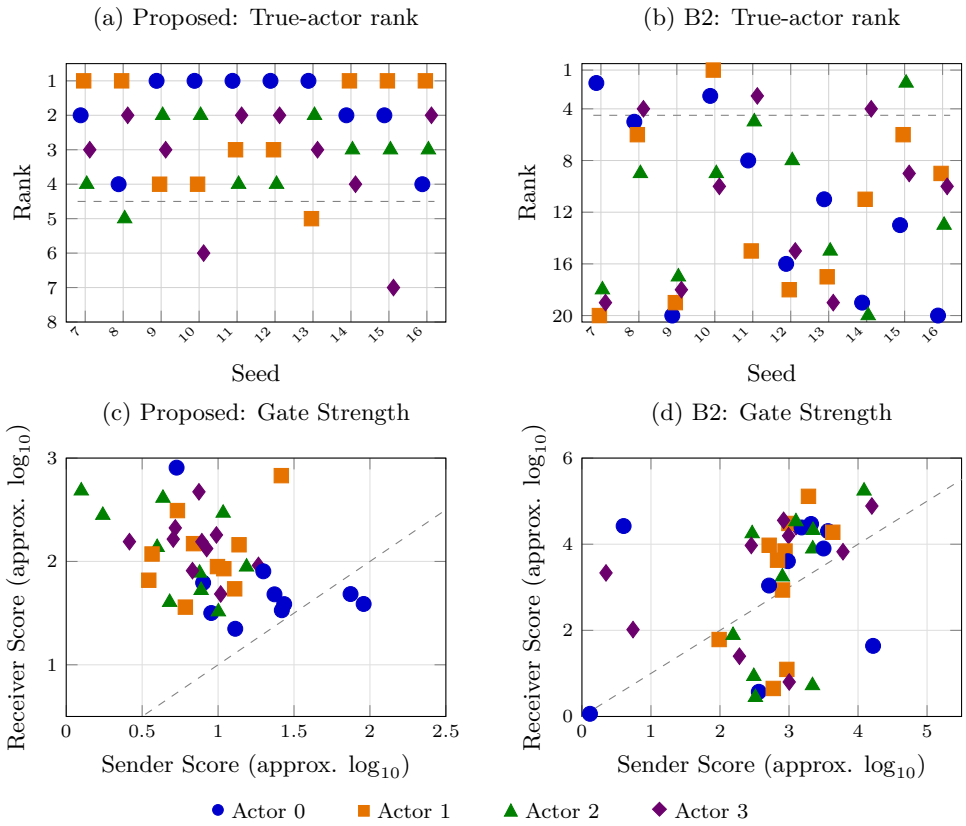

Figure 3: Diagnostic comparison of the proposed model versus the B2 baseline on the scaled benchmark. Panels (a, b) show rank diagnostics. Panels (c, d) show phase-space comparison of directional evidence. Unlike the B2 baseline, which suffers from severe rank degradation and extreme directional imbalance, the proposed model consistently maintains planted actors near the oracle top-k cutoff and demonstrates tight, stable evidence recovery across all random seeds.

CRCNS hippocampal CA1 recordings. The CRCNS hc-11 dataset contains extracellular recordings from hippocampal area CA1 collected before, during, and after novel spatial learning (Grosmark et al., 2016). We use two sessions, `Achilles_10252013` and `Cicero_09012014`. For each session, we retain the $K = 20$ most active units, use spike time as event time, and use a single mark for all spikes. The raw CA1 experiments provide label-free predictive checks on real neural event streams, while the CA1 planted-assembly experiments add a known four-unit episodic assembly to the real background. These planted assemblies are additive rather than volume-neutral, so we treat them as a real-background replication of the exposure pattern rather than as the cleanest theorem-aligned exposure test. Their purpose is to ask whether planted transient assemblies can be recovered from realistic neural burst structure, and whether modular temporal segmentation without joint actor assignment fails as predicted.

FTX-collapse trade-event background. The financial testbed uses public Binance aggregate-trade data from the November 2022 FTX-collapse period (Binance, 2022; Vidal-Tomás et al., 2023). We process USD-M futures aggregate trades for 20 crypto assets from Nov. 6–11, 2022, filter to the largest 1% of notional trades within each symbol, map assets to point-process actors, and use taker-side direction as the mark: buyer-initiated flow versus seller-initiated flow. We then analyze rebased two-hour windows around the collapse, with the Nov. 8 16:00–18:00 UTC window used as the main peak-stress background. The FTX processing step plants an FTT-rooted receiver cascade into selected assets and evaluates whether the model recovers the source, receiver targets, and active mark-channel lift. This is not a causal analysis of the FTX collapse; it is a marked-channel stress test on a highly nonstationary real market background, designed to

check whether the active block can recover directional cross-asset propagation that an unmarked model would collapse into marginal volume.

6.2 How to read the evidence

The empirical section triangulates the same claim from multiple directions. ❶ Synthetic known-truth experiments test recovery when the identifiable active-kernel structure in Theorem 1 is present and the exposure conditions in Theorem 3 are favorable. ❷ Controlled ablations test collapse when a required condition is removed, including the birth-state-blind collapse predicted by Theorem 2. ❸ Posterior-accuracy stress tests examine whether Assumption 2, item 4 (Condition C6 in Appendix A.3) is doing real work for fine directional recovery. ❹ The UCDP GED volume-neutral planted-coordination experiment tests the N_{eff} exposure logic of Theorem 3 on a real social-science background. ❺ The CRCNS CA1 planted-assembly experiments replicate the exposure pattern on a real neural event background. ❻ Raw CA1 and FTX predictive checks are reported without planted labels and are interpreted only as label-free plausibility diagnostics. ❼ The FTX planted receiver cascade and mark-channel recovery experiments stress-test the directed and marked active block in Theorem 1 and the actor–mark separation issue in Proposition 1.

The intended interpretation is not that every dataset verifies every theorem condition. Each layer has a specific evidential role: synthetic data support the theorem-clean claims, real-background planted experiments test robustness to realistic burstiness, raw real-data fits support predictive plausibility, and ablations show that the estimator's components correspond to identifiable failure modes rather than arbitrary modeling choices.

6.3 Baselines and design ablations

We compare the proposed joint switching Hawkes estimator against four simpler alternatives. B1 removes the latent regime field by fitting a static Hawkes model with gates; it tests whether learning *when* coordination is active matters beyond learning *who* is connected. B2 uses a modular HMM-plus-Hawkes pipeline; it tests whether segmenting active intervals separately from network estimation is sufficient, or whether state and network assignment must be inferred jointly. B3 fits a static Hawkes model followed by spectral recovery; it tests whether time-averaging destroys the short-lived coordination signal. B4 is an external change-point Hawkes pipeline; it tests whether change-point segmentation followed by static Hawkes fitting can recover the active subgroup without birth-time gated excitation, sender/receiver gates, or joint activation–network assignment.

We also include four identifiability ablations that deliberately remove conditions required by the theory. D1 replaces birth-time gating with present-time gating, testing whether excitation must be attributed to the parent event's birth regime rather than to the regime active when the effect is evaluated. D2 collapses separate sender and receiver gates into a single actor gate, testing whether role separation is necessary for identifying initiators and receivers, especially hubs. D3 replaces the low-rank active coupling block with a dense parameterization, testing whether a parsimonious active interaction field is needed for stable recovery rather than unrestricted pairwise active couplings. D4 enforces $\beta_0 = \beta_1$ and $\rho_0 = \rho_1$, removing temporal-shape and mark-channel separation; this probes the birth-state-blind collapse described by Theorem 2.

6.4 P1: Active exposure controls recovery

The exposure prediction is clearest in the synthetic benchmarks, where the active path and planted subgroup are known. Recovery is bimodal across seeds, and the split is controlled by realized active-born subgroup exposure: on the base benchmark, 13/20 seeds reach F1 = 1, while partial-recovery seeds average 132 active-born subgroup events versus 188 for perfect-recovery seeds. Increasing the horizon in the scaled benchmark raises perfect recovery from 13/20 to 18/20, while preserving high recovery at larger K. Table 3 gives the benchmark-level member-set recovery, and Figure 3 shows that the proposed model keeps planted actors near the oracle top-k cutoff more stably than the modular baseline.The sparse directed-edge diagnostic shows that the same active score also contains recoverable ordered-edge information, although edge-level recovery is harder than member-set or hub recovery. In a powered synthetic benchmark with a larger active subgroup,

Table 5: Recovery on real backgrounds with additively injected coordination, swept over data-generating injection strength s_{inj}. **Top:** four-actor active member-set recovery on real CA1 (`Achilles_10252013`; $K = 20$, 5,000 background events, 20 seeds); entries are mean top-k F1 with $k = 4$ (precision=recall=F1). **Bottom:** targeted receiver-cascade recovery on the real Binance FTX peak window (7 seeds): an FTT-rooted cascade is injected into `BNBUSDT`, `AVAXUSDT`, `UNIUSDT`. "Target recv. F1" = whether the three planted targets occupy the top-3 non-FTT receiver scores; "FTT in top-2" = whether the injected source stays among the top-2 active members. Higher is better throughout. The FTX panel has no baseline rows because it is a targeted recovery probe against a background that already carries an endogenous FTT receiver role.

	Injection strength s_{inj}					
Method / target	1.0	1.5	2.0	3.0	4.0	Mean
CA1 (Achilles): subgroup member-set F1, $k = 4$						
Proposed (ours)	**0.58**	**0.75**	**0.75**	**0.67**	**0.75**	**0.70**
B1: static-gate Hawkes	0.25	0.25	0.25	0.25	0.25	0.25
B2: modular HMM	0.00	0.00	0.00	0.00	0.00	0.00
B3: static SVD Hawkes	0.17	0.17	0.25	0.42	0.42	0.28
FTX peak (Binance): receiver-target recovery (proposed model)						
Target recv. F1	0.133	0.200	0.400	0.800	0.933	—
FTT in top-2 active	1.000	1.000	1.000	1.000	1.000	—

$|S^\star| = 10$, we planted a non-dense directed adjacency inside the subgroup with 18 true ordered edges among 90 ordered non-self candidate pairs. Ranking pairs by the gauge-invariant score $\widehat{s}_{j\to i} = \widehat{g}_j \widehat{h}_i \,\mathrm{softplus}(\widehat{\mathbf{u}}_j^\top \widehat{\mathbf{v}}_i)$, the proposed model achieved F1@$|E^\star| = 0.508 \pm 0.114$ over 20 independently simulated datasets, compared with the random top-$|E^\star|$ chance baseline $18/90 = 0.200$. Directed-edge AUC was 0.729 ± 0.045, above the chance value of 0.5. Thus the recovered active score carries ordered-edge information beyond membership alone, but the gap between edge-level and member-level recovery confirms that directed-edge recovery is the harder task.

The UCDP GED experiment is the main real-background exposure test because the planted coordination is volume-neutral. The original planted-member background count is fixed across the sweep, while realized planted active-born exposure increases from 25 to 200. As N_{eff} grows, the theorem-scale score $\lambda_T = \sqrt{\log(K^2/0.05)/N_{\text{eff}}}$ decreases from 0.535 to 0.189. The static-gate baseline remains near the 4/8 chance-scale member-F1 floor, whereas the proposed estimator improves from 0.525 to 0.725 member F1 and from 0.483 to 0.683 receiver F1. At the highest exposure level, paired comparisons show a +0.250 member-F1 improvement over B1 (95% CI $[+0.175, +0.325]$) and a +0.217 receiver-F1 improvement (95% CI $[+0.117, +0.317]$). We therefore treat UCDP as direct real-background support for (P1), with exact support recovery reported as directional but underpowered in this run (Table 6).

The CRCNS CA1 planted assemblies provide a complementary real-background exposure diagnostic. The completed CA1 planted-assembly design is additive rather than volume-neutral, so we do not treat it as the primary theorem-aligned real-background test. Nevertheless, the baseline behavior makes the interpretation useful: B1 remains near the 4/20 chance-scale floor, B2 remains at 0.00 member F1, and B3 rises only slowly, while the proposed model reaches mean F1 0.70. At the seed level, realized injected active-born counts order the recovery outcomes; for example, seeds with 306/278/211 injected events produce 1.0/0.75/0.5 recovery. This is the CA1 analogue of (P1): recovery follows realized active evidence more than nominal injection strength (Table 5).

The FTX planted receiver cascade is not a clean (P1) theorem test, but it is consistent with the same dose-response logic. Target receiver recovery rises from 0.133 to 0.933 as the planted receiver cascade strengthens, and the injected FTT source remains among the top-two active members throughout (Table 5). Because the financial background is highly nonstationary and already contains endogenous stress structure, we use this only as cross-domain evidence that the marked active block can recover a planted channel on an unfamiliar real background.

Table 6: UCDP GED volume-neutral exposure-threshold on India (2007–2011), $K = 8$, $k = 4$, 20 seeds. Exposure is realized planted active-born subgroup events N_{eff}, not nominal injection strength. The original planted-member background count is fixed across the sweep (49.4 mean minimum count), while λ_T decreases with exposure.

Method	N_{eff}	λ_T	Member F1	Receiver F1	Active AUC
B1 static gates	25	0.535	0.500	0.467	—
B1 static gates	50	0.378	0.513	0.483	—
B1 static gates	100	0.267	0.500	0.483	—
B1 static gates	150	0.218	0.513	0.467	—
B1 static gates	200	0.189	0.475	0.467	—
Proposed	25	0.535	0.525	0.483	0.714
Proposed	50	0.378	0.588	0.500	0.822
Proposed	100	0.267	0.638	0.533	0.935
Proposed	150	0.218	0.700	0.617	0.953
Proposed	200	0.189	0.725	0.683	0.957

Table 7: Synthetic true-$\alpha_{\min}$ weakest-link sweep for P2. The active block is a hub-to-receiver star with fixed aggregate receiver-directed active mass. Each cell uses 20 seeds. Recovery is the fraction of seeds in which the weak receiver is recovered; the threshold is the smallest $\pi_1 T$ at which recovery reaches 0.70. Positive weak-link levels that do not cross within the swept range are right-censored.

$\alpha_{\min}/\alpha_{\text{unif}}$	Weak count @ $\pi_1 T = 60$	Weak count @ $\pi_1 T = 480$	Rec.@60	Rec.@120	Rec.@240	Rec.@480
0.00	4.60	33.95	0.05	0.15	0.10	0.10
0.25	9.00	70.10	0.20	0.55	0.50	0.65
0.35	11.15	85.20	0.30	0.60	0.50	0.60
0.50	13.65	103.90	0.40	0.55	0.85	0.65

6.5 P2: Weakest-link scaling is a sharp threshold prediction

(P2) tests a different coordinate from the exposure result in (P1). It asks whether, at fixed aggregate active-event rate, the active exposure time required to recover the weakest link scales as $1/\alpha_{\min}^2$. The synthetic protocol therefore designates one receiver as the weak link, varies its true active coupling $\alpha_{\min}$ in the generative model, redistributes the remaining active mass to the other receivers so the aggregate receiver-directed active mass stays fixed, and sweeps $\pi_1 T$. We define the empirical threshold as the smallest active duration at which the weak receiver is recovered in at least 70% of seeds. If at least three positive weak-link levels have uncensored thresholds, the fitted slope of $\log(\pi_1 T_{\text{thresh}})$ versus $\log(\alpha_{\min})$ estimates the empirical exponent; the theorem predicts -2.

The quadratic threshold prediction is tested on a synthetic background where the weak receiver has no planted weak-link advantage beyond its background floor and the true $\alpha_{\min}$ is set by the generative model. At fixed aggregate active rate, the EM estimator requires longer active sojourns to recover the weak link as $\alpha_{\min}$ decreases, consistent with the threshold ordering in Theorem 3.[1]

Table 7 reports the synthetic weakest-link sweep. The run verifies the intended theorem coordinate. Realized exposure increases with horizon, with mean N_{eff} rising from 150.85 at $\pi_1 T = 60$ to 1192.75 at $\pi_1 T = 480$. The weak-link knob also changes the realized weak-receiver evidence while preserving the fixed aggregate design: at $\pi_1 T = 480$, the weak receiver's active-born count rises from the zero-coupling control value of 33.95 to 70.10, 85.20, and 103.90 as $\alpha_{\min}/\alpha_{\text{unif}}$ increases from 0.25 to 0.35 and 0.50. Thus the experiment varies the true weakest coupling while keeping the aggregate receiver-directed active mass fixed.

[1] In the present grid with a 70% recovery criterion, only $\alpha_{\min}/\alpha_{\text{unif}} = 0.50$ crosses within the swept range $\pi_1 T \leq 480$; the weaker positive levels are right-censored. Therefore the fitted log–log slope and confidence interval are not reported for this run. A wider grid that uncensors at least three positive weak-link levels would report the fitted slope and whether its confidence interval covers -2.

The threshold pattern in Table 7 is right-censored. The strongest positive weak-link level, $\alpha_{\min}/\alpha_{\text{unif}} = 0.50$, crosses the 70% recovery threshold at $\pi_1 T = 240$. The weaker positive levels do not cross within the swept range: their maximum recovery rates are 0.65 for $\alpha_{\min}/\alpha_{\text{unif}} = 0.25$ and 0.60 for $\alpha_{\min}/\alpha_{\text{unif}} = 0.35$. The zero-coupling control remains near the background false-recovery floor, with recovery rates between 0.05 and 0.15 across the same horizons. We therefore report (P2) as a censored weakest-link threshold diagnostic: the synthetic experiment validates the true-$\alpha_{\min}$ coordinate and the expected difficulty ordering, but it does not provide enough uncensored threshold points to fit a reliable log–log slope or claim an empirical -2 exponent.

The UCDP weak-link sweep is retained as a real-background diagnostic rather than as evidence for (P2). It varies a weak-link-like participation share on a volume-neutral social-science background, but the resulting threshold curve does not instantiate the sharp quadratic boundary. CA1 planted assemblies, raw CA1, and FTX are not used for (P2): CA1 varies injected exposure but not a controlled weakest link, raw CA1 has no planted weak link, and FTX is a targeted receiver-cascade and marked-channel stress test rather than a controlled $\alpha_{\min}$ sweep.

6.6 P3: Posterior accuracy is the bridge from oracle recovery to the implemented estimator

The posterior-accuracy condition is most cleanly tested synthetically, where the simulator exposes the true parameters and latent regime path. For each seed we compute the true-parameter oracle E-step posterior $\gamma_b^\star(1)$, align the fitted posterior $\widehat{\gamma}_b(1)$ by one global state-label orientation, and measure the event-weighted gap

$$\overline{\Delta}_\gamma = \frac{\sum_b N_b |\widehat{\gamma}_b(1) - \gamma_b^\star(1)|}{\sum_b N_b}, \qquad R_\gamma = \overline{\Delta}_\gamma / \lambda_T, \qquad \lambda_T = C_0 \sqrt{\frac{\log(K^2/\delta)}{N_{\text{eff}}}}.$$

Across 20 seeds, the natural generalized-EM fit has $\overline{\Delta}_\gamma \approx 0.089$ and $R_\gamma \approx 0.53$, while maintaining high top-k member recovery. A controlled posterior-corruption stress test then raises R_γ monotonically; directional recovery degrades once $R_\gamma \gtrsim 1.2$, and temporal localization falls to chance at full corruption. Swapping the fitted posterior for the oracle posterior on identical data raises hub accuracy from approximately 0.60 to 0.85. These results support the role of the posterior-accuracy assumption (Assumption 2, item 4; Condition C6 in Appendix A.3) as the bridge between the oracle support-recovery argument and the implemented generalized-EM estimator.

For real-background planted coordination, R_γ can be measured only relative to the planted active-window path, not to a true-parameter smoothing posterior for the real background. We therefore use UCDP and CA1 only as hard-path diagnostics, not as theorem-clean posterior-accuracy tests. In UCDP, pooled R_γ bands are not interpreted across exposure levels because $R_\gamma = \overline{\Delta}_\gamma/\lambda_T$ changes with N_{eff}; pooling would confound posterior error with the shrinking theorem scale. In CA1, the planted windows similarly provide a useful posterior diagnostic, but raw CA1 remains conditional on the unverified posterior-accuracy assumption.

The raw CA1 and FTX fits have no planted active path or oracle posterior, so we do not present them as (P3) tests. Instead, Table 8 is a label-free predictive diagnostic: the proposed model has the best comparable held-out likelihood on both CA1 sessions and on the FTX peak window, while the estimated active and dormant decay scales separate on CA1. These observations support the plausibility of useful latent structure, but they do not verify the posterior-accuracy condition required by Theorem 3.

6.7 P4: Removing separability produces the predicted collapse modes

(P4) is tested primarily through synthetic ablations because they isolate which piece of the identifiable structure is removed. D1 replaces birth-time gating with present-time gating, D2 collapses sender and receiver roles into a single gate, D3 relaxes the low-rank active block, and D4 shares temporal decay and mark transitions across regimes. Figure 4 shows that member-set F1 can saturate and stop discriminating at scale, while hub accuracy exposes the design failures. Present-time gating collapses hub accuracy to 0.35/0.20, and the single-gate ablation remains below the full model even when membership F1 is high. The D4 shared-β, ρ ablation gives the most direct kernel-level collapse diagnostic. In this variant, the active and dormant temporal scales and mark-transition channels are forced to coincide, so the model has no temporal-shape or mark-channel route for separating routine from active excitation. As intended, the fitted separation quantities collapse to zero:

Table 8: Held-out predictive fit on three real event backgrounds: two raw hippocampal CA1 sessions from CRCNS hc-11 (`Achilles_10252013`, `Cicero_09012014`; $K = 20$ units, $M = 1$, contiguous 80/20 split, completed paired predictive runs with $n = 14$ seeds) and the Binance USD-M FTX-collapse peak window (Nov. 8 16:00–18:00 UTC; $K = 20$ assets, 61,817 filtered events, $n = 7$ seeds). Held-out NLL/event is the latent-CTMC-marginal interval log-likelihood on the test window; lower is better. The proposed model, B1, and B3 share the marked-event likelihood and are directly comparable; B2 reports the interval-count likelihood of its modular HMM segmentation stage (shown for completeness). Standard deviations are not reported across seeds that have between-seed variation below table precision, which occurs in near-deterministic fixed-split fits. ACT. is the mean posterior active occupancy.

	Achilles (CA1)		Cicero (CA1)		FTX peak	
Method	NLL/event ↓	ACT.	NLL/event ↓	ACT.	NLL/event ↓	ACT.
Marked-event held-out likelihood (comparable)						
Proposed (ours)	**−1.46** (0.04)	0.88	**−1.938**	0.93	**−2.041** (0.003)	0.40
B1: static-gate Hawkes	−1.340	—	−1.799	—	−2.035 (0.003)	—
B3: static SVD Hawkes	−1.130	—	−1.692	—	+0.391	—
Interval-count HMM diagnostic (not comparable)						
B2: modular HMM	+0.030	0.58	+0.208	0.62	+0.071	0.72
Δ(B1−Proposed)	0.12		0.139		0.010	

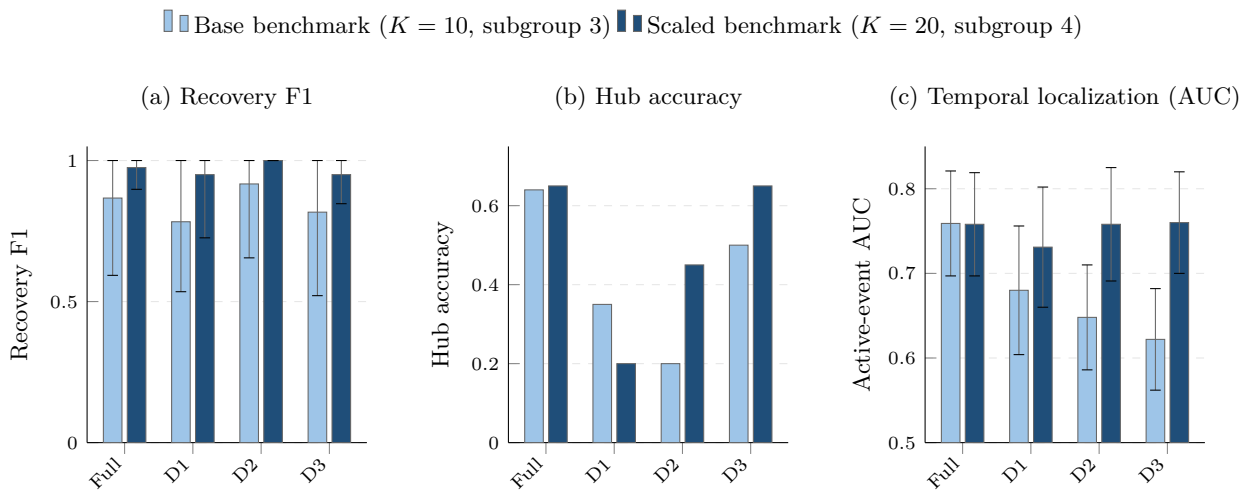

Figure 4: Identifiability ablations on the base and scaled synthetic benchmarks. (a) Subgroup recovery F1 saturates and stops discriminating in the scaled benchmark. (b) Hub accuracy more clearly exposes design tradeoffs. (c) Temporal localization (AUC). Error bars are shown for F1 and AUC; hub accuracy is shown without error bars for visual clarity. The final column reports the stronger shared-β, ρ collapse for the scaled benchmark, enforcing both $\beta_0 = \beta_1$ and $\rho_0 = \rho_1$.

$|\widehat{\beta}_1 - \widehat{\beta}_0| = 0.000 \pm 0.000$ and $\|\widehat{\rho}_1 - \widehat{\rho}_0\|_F = 0.000 \pm 0.000$. Support ranking may still look plausible, but active-kernel reconstruction fails: the recovered $\widehat{A}_0/\widehat{A}_1$ cosine is only 0.348 ± 0.120, and the active kernel has relative error 0.863 ± 0.107. This is the practical version of the collapse predicted by Theorem 2: once temporal and mark-channel separation are removed, a model can preserve some coarse actor ranking while losing the active excitation block itself.

The CA1 planted assemblies give the clearest real-background (P4) example: segmentation without assignment. The modular HMM+Hawkes baseline can locate active intervals, but its member-set F1 remains 0.00 at every injection strength (Table 5). Thus B2 can solve part of the temporal problem while failing the joint

Table 9: FTX receiver-target cascade: learned active mark-channel lift $\rho^{(1)}_{m'm} - \rho^{(0)}_{m'm}$ (mark 0 = buyer-initiated, 1 = seller-initiated; mean over 5 seeds). Row-stochasticity forces the diagonal lifts to be the negatives of the off-diagonals, so the two cross-mark entries summarize the channel. The seller-row sell-then-buy lift ($1 \to 0$) grows monotonically with injection strength while the buyer-row lift ($0 \to 1$) is near-constant, evidencing a directional cross-asset sell-to-buy absorption channel.

Injection strength s_{inj}	1.0	1.5	2.0	3.0	4.0
$0 \to 1$ lift (buy-then-sell)	+0.037	+0.027	+0.033	+0.035	+0.051
$1 \to 0$ lift (sell-then-buy)	+0.026	+0.053	+0.101	+0.190	+0.254

assignment problem. This matches the identifiability framing: detecting an active interval is not enough when the active actors and roles must be recovered jointly with the latent regime.

Raw CA1 likelihoods and the FTX stress test provide additional separation diagnostics. On CA1, the proposed model improves held-out likelihood over the static-gate and static-SVD baselines and separates dormant from active decay scales (Table 8). On FTX, the marked active block recovers a seller-row sell-then-buy lift that grows monotonically with injection strength (Table 9). This mark-channel result is not a causal claim about markets; it is evidence that the marked factorization can recover a planted cross-asset channel that an unmarked model would collapse.

Together, (P1)–(P4) give a coherent empirical picture. Active coordination is recoverable when realized active exposure is sufficient, the posterior bridge is accurate enough, and birth-time/role/mark separability is preserved. When those links are broken, a method may still fit likelihood, segment active periods, or rank active members, while missing the coordinating roles, active kernel, or marked propagation channel.

7 Conclusion

We proposed a switching marked point-process formulation for episodic coordination, treating recovery as a coupled identifiability and estimation problem rather than scalar anomaly detection. Recovery requires effective active exposure, dormant–active separation, birth-time attribution of excitation, and an anchor separating actor couplings from mark-transition weights; our birth-time gated estimator is built around these requirements, which turns each identifiability condition into a concrete empirical test that should degrade when the condition is removed.

Across synthetic, UCDP, and CRCNS experiments, recovery is most consistent in the theorem's own coordinates: realized active-born exposure N_{eff}, weakest-link strength $\alpha_{\min}$ at fixed aggregate active mass, posterior accuracy R_γ, and exact or near-exact support recovery. Synthetic ablations supply the clearest collapse tests, UCDP the cleanest volume-neutral real-background exposure test, and CRCNS a neural real-background replication of the exposure pattern. The dedicated synthetic weakest-link sweep varies the true $\alpha_{\min}$ coordinate directly while holding aggregate receiver-directed active mass fixed, and weaker links require longer active sojourns to recover; in the present grid the threshold curve is right-censored, so we report the observed ordering as a controlled diagnostic consistent with the theory rather than claiming an empirical -2 slope. When timing, excitation, marks, and directional role structure are preserved, short-lived subgroups are recoverable even on noisy real backgrounds; when those links are broken, a model may still segment time or fit likelihood while missing the coordinating actors, roles, and channels.

8 Limitations

The latent-state half of Theorem 1 and all of Theorem 3 are conditional on posterior accuracy. In particular, Theorem 3 assumes the posterior-weighted E-step stays close to the oracle birth-state posterior at $o(\lambda_T)$ scale; it is not a global EM convergence guarantee and does not prove this posterior concentration from primitive conditions. When active exposure is small, kernels are near-collinear, or emissions are weakly separated, the posterior may not concentrate. Informative priors on active occupancy can regularize the activation

timeline but do not create identifiability when the structural separation conditions fail. In controlled synthetic experiments, the posterior-accuracy stress test compares the fitted E-step posterior to a true-parameter oracle E-step posterior and shows that the natural EM fit has small event-weighted posterior error, while controlled posterior corruption increases the normalized posterior gap associated with Assumption 2, item 4 (Condition C6 in Appendix A.3). This supports the operative role of Condition C6, but does not prove the theorem's uniform $o(\lambda_T)$ condition or its unknown constants. Raw CA1 and FTX experiments have no oracle birth-state posterior, so those raw-data fits remain conditional on local E-step posterior accuracy.

The synthetic weakest-link experiment confirms the difficulty ordering predicted by Theorem 3 but, in the present exposure grid, only the strongest positive level uncenses, so we do not fit the log–log exponent; a wider grid that uncenses ≥ 3 levels is required to estimate it. The oracle restricted-eigenvalue condition is also a local regularity assumption on a nonconvex gated low-rank objective, not a primitive consequence of the model; if the oracle Hessian is nearly flat or ill-conditioned along gate/embedding directions, the support-recovery theorem does not apply. Recovery is exposure-bounded—the share of perfectly recovered seeds rises with horizon ($13/20 \to 18/20$) and real-CA1 planted coordination F1 is ranked by realized active-event count—so deployment with faint weakest links needs longer windows or auxiliary side information. A further empirical limitation is that the external B4 comparison is a lightweight two-step change-point Hawkes baseline, not a full neural change-point temporal point-process implementation. It is included to test whether a concrete change-point-plus-Hawkes pipeline suffices on the synthetic recovery task, but it should not be read as an exhaustive benchmark over all change-point Hawkes variants. Finally, the identifiability guarantees are specific to the binary latent field; multi-state coordination is natural to add at the modeling level but is not covered here.

References

Emmanuel Bacry, Khalil Dayri, and Jean-François Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *European Physical Journal B*, 85:157, 2012.

Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

Emmanuel Bacry, Martin Bompaire, Stéphane Gaïffas, and Jean-François Muzy. Sparse and low-rank multivariate hawkes processes. *Journal of Machine Learning Research*, 21(50):1–32, 2020.

Peter J. Bickel, Ya'acov Ritov, and Alexandre B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

Binance. Binance public data. `https://data.binance.vision/`, 2022. Aggregate trade data archive.

Pierre Brémaud. *Point Process Calculus in Time and Space*. Springer, Cham, 2020.

Peter Bühlmann and Sara van de Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, 2011.

Ido Cohn, Tal El-Hay, Nir Friedman, and Raz Kupferman. Mean field variational approximation for continuous-time bayesian networks. *Journal of Machine Learning Research*, 11:2745–2783, 2010.

D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer, New York, 2003.

Andres D Grosmark, John Long, and György Buzsáki. Recordings from hippocampal area CA1, PRE, during and POST novel spatial learning. *CRCNS.org*, 2016. doi: 10.6080/K0862DC5.

Niels Richard Hansen, Patricia Reynaud-Bouret, and Vincent Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 2015.

Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1): 83–90, 1971.

Pritam Koley, Hamza Alimi, Sahil Singla, Saptarshi Bhattacharya, Niloy Ganguly, and Abir De. Differentiable change-point detection with temporal point processes. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, pp. 6940–6955. PMLR, 2023.

Pascal Lezaud. Chernoff-type bound for finite markov chains. *Annals of Applied Probability*, 8(3):849–867, 1998.

Scott Linderman and Ryan Adams. Discovering latent network structure in point process data. In *Proceedings of the International Conference on Machine Learning*, 2014.

Hongyuan Mei and Jason M. Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*, 2017.

Sahand Negahban, Bin Yu, Martin J. Wainwright, and Pradeep Ravikumar. A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. In *Advances in Neural Information Processing Systems*, 2009.

Patricia Reynaud-Bouret. Compensator and exponential inequalities for some suprema of counting processes. *Statistics & Probability Letters*, 62(1):49–61, 2003.

Magnus Öberg Shawn Davies, Therése Pettersson. Organized violence 1989–2025, and violent political protests. *Journal of Peace Research*, 50(4), 2026. doi: 10.1093/jopres/xjag046.

Ralph Sundberg and Erik Melander. Introducing the UCDP georeferenced event dataset. *Journal of Peace Research*, 50(4), 2013.

David Vidal-Tomás, Antonio Briola, and Tomaso Aste. Ftx's downfall and binance's consolidation: The fragility of centralised digital finance. *Physica A: Statistical Mechanics and its Applications*, 625:129044, 2023. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2023.129044. URL `https://www.sciencedirect.com/science/article/pii/S037843712300599X`.

Martin J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, 2009.

Ting Wang, Mark Bebbington, and David Harte. Markov-modulated hawkes process with stepwise decay. *Annals of the Institute of Statistical Mathematics*, 64(3):521–544, 2012.

Jue Wu, Olivia G. Ward, James Curley, and Tian Zheng. Markov-modulated hawkes processes for sporadic and bursty event occurrences in social interactions. *The Annals of Applied Statistics*, 16(2):1171–1190, 2022.

Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. Learning granger causality for hawkes processes. In *Proceedings of the International Conference on Machine Learning*, 2016.

Ziyu Zhang, Xiaoyu Lu, and Feng Zhou. Conjugate bayesian two-step change point detection for hawkes process. In *Advances in Neural Information Processing Systems*, 2024.

Ke Zhou, Hongyuan Zha, and Le Song. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2013a.

Ke Zhou, Hongyuan Zha, and Le Song. Learning triggering kernels for multi-dimensional hawkes processes. In *Proceedings of the International Conference on Machine Learning*, 2013b.

Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. Transformer hawkes process. In *Proceedings of the International Conference on Machine Learning*, 2020.

A Proofs

A.1 Theorem 1: Identifiability

Proof. The proof proceeds as follows. We augment the observed process with a latent birth-state label and condition on it to obtain a labeled marked Poisson cluster representation; identify the composite triggering kernel from the first-order branching (resolvent) structure of that representation; separate the routine and active blocks using linear independence of the temporal kernels; and finally resolve the actor–mark scaling ambiguity by normalization.

Step 1. Augmented Hawkes representation. Let $Z(t) \in \{0, 1\}$ denote the latent CTMC governing the regime at time t, with transition rates η_{on} (from dormant to active) and η_{off} (from active to dormant). By assumption (a), $\eta_{\text{on}} > 0$ and $\eta_{\text{off}} > 0$, so both states are recurrent and the stationary distribution $\pi = (\pi_0, \pi_1)$ satisfies $\pi_r > 0$ for each $r \in \{0, 1\}$.

Augment each observed event (t_n, i_n, m_n) with its *birth-state label* $r_n := Z(t_n^-) \in \{0, 1\}$, indicating the regime active at the instant just before the event. On the expanded type space (i, m, r), $i \in [K]$, $m \in \mathcal{M}$, $r \in \{0, 1\}$, the process is a linear marked Hawkes process with conditional intensity

$$\lambda_{(i,m,r)}(t) = \mu_i^{(r)} \mathbf{1}[Z(t^-) = r] + \sum_{r'} \sum_{j,m'} \alpha_{j\to i}^{(r')} \rho_{m'm}^{(r')} \int_0^{t^-} \kappa_{r'}(t - s)\, dN_{(j,m',r')}(s),$$

where $N_{(j,m',r')}$ denotes the counting process restricted to events of expanded type (j, m', r'). The regime-specific excitation matrices are

$$B_r\big[(j, m'), (i, m)\big] := \alpha_{j\to i}^{(r)} \rho_{m'm}^{(r)}, \qquad r \in \{0, 1\}.$$

The observed (unlabeled) process is the marginal obtained by summing over r: $N_{(i,m)}(t) = N_{(i,m,0)}(t) + N_{(i,m,1)}(t)$.

Step 1b. Conditioning and the labeled cluster representation. Because $Z(\cdot)$ is an exogenous latent modulator, the *unlabeled* process is not a stationary linear Hawkes process: its second-order structure carries an extra contribution from fluctuations of Z. We therefore argue identifiability on the birth-state–*labeled* process, which is observed directly under the first hypothesis of the theorem and, under the second hypothesis, is recovered with error that vanishes uniformly over event-containing intervals. Conditional on the realized regime path $Z(\cdot)$, an event of expanded type (j, m', r') excites future intensity only through the fixed matrix $B_{r'}\kappa_{r'}$, so the labeled process admits the marked Poisson cluster (branching) representation of a linear Hawkes process with regime-stamped, time-invariant triggering matrices $\{B_0, B_1\}$. Identifiability of B_1 for the labeled law therefore implies identifiability of the active block of the observed marked law, since under either hypothesis the joint law of events and birth-state labels is determined by the data.

Step 2. First-order branching structure. For a stable linear marked Hawkes process the conditional expected offspring (Palm) kernel $\mathbf{\Psi}(\tau) := \sum_{n\geq 1} \mathbf{\Phi}^{*n}(\tau)$ is identifiable from the law of the labeled process as the dual-predictable response of the intensity to a seed event (Hawkes, 1971; Daley & Vere-Jones, 2003; Bacry et al., 2012). In the Fourier–Laplace domain it satisfies the resolvent identity

$$\mathbf{I} + \widehat{\mathbf{\Psi}}(s) = \big[\mathbf{I} - \widehat{\mathbf{\Phi}}(s)\big]^{-1}, \qquad \widehat{\mathbf{\Phi}}(s) = B_0\, \widehat{\kappa}_0(s) + B_1\, \widehat{\kappa}_1(s), \tag{12}$$

where convergence of the Neumann series is guaranteed by the stability condition in Assumption 1 ($\rho(\int_0^\infty \mathbf{\Phi}) < 1$). Assumptions (a) and (b) make both blocks nondegenerate: (a) gives $\pi_r > 0$ for $r \in \{0, 1\}$, so events are born under both regimes, and (b) gives $B_1 \neq 0$.

Step 3. Exact inversion and separation via kernel linear independence. The resolvent identity equation 12 inverts *exactly* to

$$\widehat{\mathbf{\Phi}}(s) = \mathbf{I} - \big[\mathbf{I} + \widehat{\mathbf{\Psi}}(s)\big]^{-1}. \tag{13}$$

This matrix inversion is the causal Wiener–Hopf step for stable Hawkes processes (Hawkes, 1971; Bacry et al., 2012); unlike a symmetric square root of a second-order spectrum it is unique and introduces no rotational ambiguity, because the identified object is the resolvent $[\mathbf{I} - \widehat{\mathbf{\Phi}}(s)]^{-1}$ itself. Hence $\widehat{\mathbf{\Phi}}(s)$ is identifiable as a function of s.

For the exponential family $\kappa_r(\tau) = \beta_r e^{-\beta_r \tau}\mathbf{1}[\tau > 0]$ the transforms $\widehat{\kappa}_r(s) = \beta_r/(\beta_r + s)$ are rational with distinct poles at $s = -\beta_0$ and $s = -\beta_1$. By assumption (c), $\beta_0 \neq \beta_1$, so $\widehat{\kappa}_0$ and $\widehat{\kappa}_1$ are linearly independent in $L^1(\mathbb{R}_+)$ (cf. Brémaud, 2020, Section 2.3). Each entry of $\widehat{\mathbf{\Phi}}(s)$ equals $B_0[\cdot]\,\widehat{\kappa}_0(s) + B_1[\cdot]\,\widehat{\kappa}_1(s)$, so evaluating equation 13 at two distinct points $s_1, s_2 \in \mathbb{C} \setminus \{-\beta_0, -\beta_1\}$ yields the linear system

$$\begin{pmatrix} \widehat{\kappa}_0(s_1) & \widehat{\kappa}_1(s_1) \\ \widehat{\kappa}_0(s_2) & \widehat{\kappa}_1(s_2) \end{pmatrix} \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{\Phi}}(s_1) \\ \widehat{\mathbf{\Phi}}(s_2) \end{pmatrix},$$

whose coefficient determinant

$$\widehat{\kappa}_0(s_1)\widehat{\kappa}_1(s_2) - \widehat{\kappa}_0(s_2)\widehat{\kappa}_1(s_1) = \frac{\beta_0\beta_1(\beta_0 - \beta_1)(s_1 - s_2)}{(\beta_0 + s_1)(\beta_1 + s_1)(\beta_0 + s_2)(\beta_1 + s_2)}$$

is nonzero whenever $\beta_0 \neq \beta_1$ and $s_1 \neq s_2$. Hence B_0 and B_1 are uniquely determined, and in particular the active excitation matrix B_1 is identifiable.

Remark 2. *If $\beta_0 = \beta_1$ then $\widehat{\kappa}_0 = \widehat{\kappa}_1$ and only the sum $B_0 + B_1$ is identifiable, so the active component cannot be separated from the routine component; this is why assumption (c) is necessary. For general kernels, linear independence of $\widehat{\kappa}_0, \widehat{\kappa}_1$ in $L^1(\mathbb{R}_+)$ suffices, with the two-point evaluation replaced by uniqueness of the representation in* $\operatorname{span}\{\widehat{\kappa}_0, \widehat{\kappa}_1\}$.

Step 4. Active kernel block and transfer to the observed law. Since κ_1 is known once β_1 is identified, Step 3 identifies the full active kernel block $\Phi^{(1)}_{(j,m')\to(i,m)}(\tau) = B_1[(j,m'),(i,m)]\,\kappa_1(\tau)$. Under observed birth-state labels, and likewise under a latent posterior that is consistent for the birth-state labels, the response can be stratified by the parent event's recovered birth state with vanishing label error. Dormant-born parents identify the dormant level $B_0\kappa_0$, active-born parents identify the active level $B_0\kappa_0 + B_1\kappa_1$, and the active increment $B_1\kappa_1$ is obtained by subtraction. This birth-state route does not require condition (c). If one instead works only with a birth-state-blind composite response, then the two-kernel decomposition in Step 3 requires linear independence of κ_0 and κ_1; when the kernels are collinear and birth states are not recovered, the collapsed decomposition is not identifiable, as formalized in Theorem 2. Thus the active marked kernel block is identifiable either through recovered birth states or through temporal-shape separation.

Step 5. Separation of actor couplings from mark transitions. The active excitation matrix factorizes as

$$B_1\big[(j,m'),(i,m)\big] = \alpha^{(1)}_{j\to i}\,\rho^{(1)}_{m'm}.$$

This is a rank-structured factorization: B_1 is the Kronecker (tensor) product of the actor coupling matrix $A^{(1)} = [\alpha^{(1)}_{j\to i}]_{i,j\in[K]}$ and the mark-transition matrix $R^{(1)} = [\rho^{(1)}_{m'm}]_{m',m\in\mathcal{M}}$: $B_1 = A^{(1)} \otimes R^{(1)}$. Without further constraints, the Kronecker factorization $B_1 = A \otimes R$ admits a one-parameter family of solutions $(cA) \otimes (c^{-1}R)$ for any $c \neq 0$, so $A^{(1)}$ and $R^{(1)}$ are individually identifiable only up to a shared scalar.

If $\rho^{(1)}_{m'm}$ is known, then $R^{(1)}$ is fixed and $A^{(1)}$ is uniquely determined by the relation $\alpha^{(1)}_{j\to i} = B_1[(j,m'),(i,m)]/\rho^{(1)}_{m'm}$ for any (m',m) with $\rho^{(1)}_{m'm} > 0$. If instead the row-stochastic normalization

$$\sum_{m\in\mathcal{M}} \rho^{(1)}_{m'm} = 1 \qquad \text{for each } m' \in \mathcal{M}$$

is imposed, then $c = 1$ is the unique scalar preserving this constraint, so the factorization is again unique. In either case, $\alpha^{(1)}_{j\to i}$ is identifiable for all $j, i \in [K]$.

This completes the proof. □

Remark 3 (Unobserved latent state)**.** *When the birth-state labels are not observed, identification in Theorem 1 proceeds through its second hypothesis (a consistent regime posterior), which in turn requires the switching process to generate distinguishable interval emissions: if the dormant and active emissions coincide, the birth-state posterior cannot concentrate and the labeled cluster representation of Step 1b is not recovered. Finite-sample recovery then depends on posterior concentration, made precise in Appendix A.3 by Condition C6 and Remark 4 for Theorem 3.*

A.2 Theorem 2: Collinearity and birth-state-blind collapse

Proof. Because both kernels are normalized, collinearity implies equality. Indeed, if $\kappa_0 = c\kappa_1$ and $\int_0^\infty \kappa_r(\tau)\,d\tau = 1$ for $r \in \{0,1\}$, then $1 = c$, and hence $\kappa_0 = \kappa_1 =: \kappa$. Substituting this into Eq. 1 gives

$$\lambda_i(t, m \mid \mathcal{H}_t, Z) = \nu_i(m) + \sum_{j,m'} \int_0^{t^-} \left[B^{(0)}_{(j,m')\to(i,m)} + Z(s) B^{(1)}_{(j,m')\to(i,m)} \right] \kappa(t-s)\, dN_j(s, m').$$

Thus the model has two regime-specific excitation levels attached to the birth state of the parent event: $B^{(0)}$ for dormant-born parents and $B^{(0)} + B^{(1)}$ for active-born parents. If the birth-state labels are observed, or if the posterior over birth states is consistent and both states have positive exposure, these two levels can be estimated separately, and the active block is then their difference. This proves the first claim and also shows why collinearity by itself is not a full-law non-identifiability result.

It remains to show what is lost under collinearity in a birth-state-blind, time-invariant single-kernel representation. Fix any $q \in (0, 1]$ and suppose the population kernel-level object available to the restricted estimator is

$$B^{\text{mix}}_q = B^{(0)} + qB^{(1)}.$$

Here q is part of the definition of the collapsed representation; it is not identified with the endogenous active-event fraction of the full switching model. We now exhibit two distinct stable structured parameterizations with the same B^{mix}_q.

Let A be any nonzero nonnegative actor-coupling matrix and let R be any row-stochastic mark-transition matrix. Set $D := A \otimes R$. Choose scalars $a_0 > 0$ and $a_1 > 0$ such that the stability condition holds, for example $(a_0 + a_1)\rho(D) < 1$. Define

$$B^{(0)} = a_0 D, \qquad B^{(1)} = a_1 D.$$

For any $\varepsilon \in (0, qa_1)$, define a second structured parameterization

$$\widetilde{B}^{(0)} = (a_0 + \varepsilon)D, \qquad \widetilde{B}^{(1)} = \left(a_1 - \frac{\varepsilon}{q}\right) D.$$

Both pairs lie in the structured family $B^{(r)} = A^{(r)} \otimes R^{(r)}$ and remain nonnegative. The admissibility bound $\varepsilon < qa_1$ ensures $\widetilde{B}^{(1)} \succeq 0$. Stability is preserved because

$$\widetilde{B}^{(0)} + \widetilde{B}^{(1)} = \left(a_0 + a_1 + \varepsilon - \frac{\varepsilon}{q}\right) D \preceq (a_0 + a_1)D,$$

where the last inequality uses $q \le 1$. Thus the active-regime branching matrix of the perturbed parameter is no larger entrywise than that of the original stable parameter. Since $\widetilde{B}^{(1)} \succeq 0$, dormant-regime stability follows from $\widetilde{B}^{(0)} \preceq \widetilde{B}^{(0)} + \widetilde{B}^{(1)}$. They have different active blocks because $\widetilde{B}^{(1)} \neq B^{(1)}$, but they induce the same collapsed kernel:

$$\widetilde{B}^{(0)} + q\widetilde{B}^{(1)} = (a_0 + \varepsilon)D + q\left(a_1 - \frac{\varepsilon}{q}\right) D = (a_0 + qa_1)D = B^{(0)} + qB^{(1)}.$$

Therefore the routine/active decomposition cannot be recovered from B^{mix}_q alone. The two structured parameterizations are generally distinguishable under the full Markov-modulated law, because their dormant

levels differ, but they are indistinguishable in the restricted collapsed single-kernel experiment. For any estimator whose input is only this collapsed kernel, the two parameterizations yield the same input while their active blocks are separated by $\|B^{(1)} - \widetilde{B}^{(1)}\| = (\varepsilon/q)\|D\|$. The usual two-point argument then gives a nonzero worst-case error lower bound over this restricted collapsed experiment, uniformly in the amount of data used to estimate B_q^{mix}. This completes the proof. □

A.3 Theorem 3: Finite-sample gate-support recovery

This appendix proves Theorem 3. The proof shows exact gate-support recovery provided the approximate E-step remains uniformly close to the oracle regime posterior after initialization. The argument deliberately uses standard ingredients: CTMC occupation concentration, counting-process tail bounds, primal-dual-witness support recovery under a restricted-eigenvalue condition, and a perturbation step from the oracle surrogate to the implemented surrogate. The model-specific contribution is the reduction of the birth-time-gated recovery problem to the effective active sample size N_{eff}, defined below.

Recovery target. Under the active-coupling parameterization

$$\alpha^{(1)}_{j\to i} = g_j h_i \,\mathrm{softplus}(\mathbf{u}_j^\top \mathbf{v}_i),$$

the sparse objects are the sender and receiver gate supports

$$S_g^\star := \{j : g_j^\star > 0\}, \qquad S_h^\star := \{i : h_i^\star > 0\}.$$

Accordingly, the recovery target is the active member set

$$S^\star := S_g^\star \cup S_h^\star.$$

This is the support notion for the present model, since once both gates are positive the pairwise active coupling is positive as well.

Effective active sample size. Let

$$S_1(T) := \int_0^T \mathbf{1}[Z(t) = 1]\, dt$$

denote the active sojourn time. Because the model uses birth-time gating, the relevant effective sample size is the number of events born in the active state from true subgroup actors:

$$N_{\mathrm{eff}} := \sum_{n=1}^{N} \mathbf{1}[Z(t_n) = 1]\, \mathbf{1}[c_n \in S^\star]. \tag{14}$$

Signal parameters. Because the factorization $\alpha^{(1)}_{j\to i} = g_j h_i \mathrm{softplus}(\mathbf{u}_j^\top \mathbf{v}_i)$ is invariant under $g \mapsto cg$ and $h \mapsto c^{-1}h$, gate magnitudes are interpreted in the penalty-balanced gauge selected by the regularizer in Eq. 4. Let $(g^{\mathrm{pb},\star}, h^{\mathrm{pb},\star})$ denote the penalty-balanced representative of the true gate orbit. We use the shorthand

$$\alpha_{\min} := \min_{j\in S_g^\star,\ i\in S_h^\star} \alpha^{(1),\star}_{j\to i}, \qquad g^{\mathrm{pb}}_{\min} := \min_{j\in S_g^\star} g_j^{\mathrm{pb},\star}, \qquad h^{\mathrm{pb}}_{\min} := \min_{i\in S_h^\star} h_i^{\mathrm{pb},\star}.$$

The first quantity is gauge-invariant. The latter two are well-defined only after fixing the penalty-balanced gauge, which is also the gauge of the penalized estimator.

Technical conditions. We work under the following conditions.

(C1) **Stability and identifiability.** The stability condition equation 1 holds and the identifiability conditions of Theorem 1 hold. The nuisance parameters $\rho^{(r)}$, β_r, and the routine block are either known, or are estimated accurately enough that their residual error is asymptotically smaller than the support-separation scale.

(C2) **Minimum signal (beta-min condition).** The active member gates, expressed in the penalty-balanced gauge, are separated from zero:

$$g_{\min}^{\mathrm{pb}} > 0, \qquad h_{\min}^{\mathrm{pb}} > 0.$$

This is the analogue of the classical beta-min condition in high-dimensional support recovery (Bühlmann & van de Geer, 2011; Wainwright, 2009). Without a signal gap between zero and nonzero coordinates, no estimator can distinguish active from inactive gates at finite sample sizes.

(C3) **Oracle active-block regularity.** If the true regime weights are used in place of the soft regime weights, then the resulting oracle active-block surrogate satisfies a local restricted-curvature condition on the active parameter block, sufficient for coordinatewise recovery of the *gate* supports. More precisely, define

$$\theta := (\{g_j\}, \{h_i\}, \{\mathrm{vec}(\mathbf{u}_j)\}, \{\mathrm{vec}(\mathbf{v}_i)\})$$

in a local gauge-fixed chart. The gate scaling gauge is fixed by the penalty-balanced representative. The embedding coordinate transformations that preserve all inner products $\mathbf{u}_j^\top \mathbf{v}_i$ are treated as a nuisance gauge and are fixed by a local orientation convention or quotiented out. Near a nondegenerate oracle point, this gauge orbit is a smooth finite-dimensional manifold, so a local slice may be chosen by the usual constant-rank/slice argument. Since the theorem targets gate supports rather than the coordinates of $\mathbf{u}_j$ and $\mathbf{v}_i$ themselves, the support object is defined over the gate coordinates:

$$S_G^\star := S_g^\star \cup S_h^\star,$$

while the embeddings associated with active members form a dense nuisance block and the embeddings of nonmembers are irrelevant once their corresponding gate is zero.
Let the combined decomposable regularizer on perturbations $\Delta = (\Delta_g, \Delta_h, \Delta_U, \Delta_V)$ be

$$\mathcal{R}(\Delta) := \lambda_g \|\Delta_g\|_1 + \lambda_h \|\Delta_h\|_1 + \lambda_e \left(\sum_{j=1}^{K} \|\Delta_{u_j}\|_2 + \sum_{i=1}^{K} \|\Delta_{v_i}\|_2 \right).$$

Let $\mathcal{M}^\star$ denote the subspace consisting of perturbations on the active gate coordinates $S_G^\star$ together with the dense embedding rows for active senders and receivers, and let $(\mathcal{M}^\star)^\perp$ denote its complement. The restricted cone is the decomposable-regularizer cone

$$\mathcal{C}_{\mathcal{R}}(S^\star) := \left\{ \Delta : \mathcal{R}\big(\Delta_{(\mathcal{M}^\star)^\perp}\big) \le 3\mathcal{R}\big(\Delta_{\mathcal{M}^\star}\big) \right\},$$

rather than a pure ℓ_1 cone over the full parameter vector. The oracle Hessian $\nabla^2 Q_{\mathrm{or}}(\theta^\star)$ satisfies

$$\frac{1}{N_{\mathrm{eff}}} \Delta^\top [-\nabla^2 Q_{\mathrm{or}}(\theta^\star)] \Delta \;\ge\; \kappa_{\mathrm{RE}} \|\Delta\|_2^2 \qquad \text{for all } \Delta \in \mathcal{C}_{\mathcal{R}}(S^\star),$$

for some constant $\kappa_{\mathrm{RE}} > 0$. This is the analogue of the restricted-eigenvalue assumptions used for M-estimators with decomposable regularizers (Negahban et al., 2009), including ℓ_1-penalized point-process models (Hansen et al., 2015). For the present model, however, it is a local oracle regularity assumption rather than a consequence of convexity: the active block is nonconvex because the gates multiply and the embeddings enter through the low-rank bilinear scores $\mathbf{u}_j^\top \mathbf{v}_i$. The condition is plausible when the oracle active-born traces provide nondegenerate variation over the active sender and receiver coordinates, the penalty-balanced gate chart removes the global scaling flat direction, the embedding gauge is fixed or quotiented out, and the routine/active traces are not nearly redundant. We do not derive primitive conditions guaranteeing this restricted-curvature property for the gated low-rank parameterization; Theorem 3 is conditional on this local curvature assumption.

(C4) **Oracle score bound.** For the oracle active sample, the gradient of the active-block surrogate is bounded in the dual norm associated with the combined regularizer $\mathcal{R}$. In particular, the gate score coordinates obey an ℓ_∞ bound and the embedding-row scores obey the corresponding group-dual ℓ_2 bound, all at order

$$\lambda_T := C_0 \sqrt{\frac{\log(K^2/\delta)}{N_{\mathrm{eff}}}}$$

with probability at least $1 - \delta/4$, for some constant $C_0 > 0$. This bound follows from exponential inequalities for suprema of martingales associated with counting processes (Reynaud-Bouret, 2003). Specifically, conditioned on the regime path $Z(\cdot)$, the score contributions from active-born events are compensated martingale increments, and the $\sqrt{\log(K^2/\delta)/N_{\text{eff}}}$ rate is obtained via a maximal inequality over the gate coordinates and embedding row groups of θ, following the approach of Hansen et al. (2015) for multivariate point process regression.

(C5) **Concentration of active exposure and effective event count.** There exist constants $c_S, c_\mu, c_N > 0$ such that

$$\Pr\left(S_1(T) \le \frac{\pi_1 T}{2}\right) \le 2e^{-c_S(\eta_{\text{on}}+\eta_{\text{off}})T}.$$

This is a Chernoff-type bound for the occupation time of a finite-state continuous-time Markov chain (Lezaud, 1998). The spectral gap of the two-state CTMC generator is $\eta_{\text{on}} + \eta_{\text{off}}$, which controls the concentration rate.

Conditional on the realized field path $Z(\cdot)$, the aggregate conditional intensity of active-born events from the true subgroup satisfies

$$\sum_{i \in S^\star} \sum_{m \in \mathcal{M}} \lambda_i(t, m \mid \mathcal{H}_t, Z(t) = 1) \ge c_\mu \bar{\mu}$$

on active intervals, where $\bar{\mu}$ is a subgroup-level aggregate active-event intensity. The corresponding effective count satisfies

$$\mathbb{E}[N_{\text{eff}} \mid Z] \ge c_\mu \bar{\mu}\, S_1(T),$$

where $\bar{\mu}$ is a lower bound on the aggregate conditional intensity of subgroup-actor events during active intervals. This convention treats $\bar{\mu}$ as the active evidence rate supplied by the subgroup as a whole; if one instead uses a per-actor lower bound $\bar{\mu}_{\text{act}}$, then $\bar{\mu} \asymp k\bar{\mu}_{\text{act}}$. The lower tail satisfies

$$\Pr\left(N_{\text{eff}} \le \frac{1}{2} c_\mu \bar{\mu}\, S_1(T) \,\middle|\, Z\right) \le e^{-c_N \bar{\mu} S_1(T)}.$$

This follows from exponential inequalities for counting processes with bounded conditional intensity (Reynaud-Bouret, 2003), see also Daley & Vere-Jones (2003), Chapter 7, for general moment and tail bounds on point process functionals.

(C6) **Local E-step posterior accuracy.** Let $\gamma_b^\star(1)$ denote the oracle posterior probability that discretized interval b belongs to the active regime. The generalized EM iterates produce approximate posteriors satisfying

$$\max_{1 \le b \le B} \left|\gamma_b^{(t)}(1) - \gamma_b^\star(1)\right| \le \varepsilon_T \qquad \text{for all sufficiently large } t,$$

with probability at least $1 - \delta/4$, where $\varepsilon_T = o(\lambda_T)$. Equivalently, the approximate surrogate Q_{approx} satisfies

$$\sup_{\Theta} |Q_{\text{approx}}(\Theta) - Q_{\text{or}}(\Theta)| \le C_Q \varepsilon_T N_{\text{eff}}$$

for some constant $C_Q > 0$.

This condition constrains the quality of the posterior approximation. It is an assumption, not a consequence of the preceding conditions. In practice it may depend on initialization, signal strength, exposure, and the separation between routine and active emissions. The mean-field variational approximation underlying our posterior-weighted trace construction (Section 5) is analogous to the variational bound for continuous-time Markov chains developed by Cohn et al. (2010), which was applied in the MMHP setting by Wu et al. (2022). We do not derive a primitive KL or mixing-rate bound for this approximation here; the support-recovery theorem therefore assumes Condition C6 directly.

Remark 4 (Primitive route to local E-step posterior accuracy)**.** *The following emission-separation and mixing requirements describe a natural route under which one would expect Condition C6 to hold with $\varepsilon_T = o(\lambda_T)$. We do not use this remark as a theorem, because a complete proof must also control the mean-field trace approximation error. Suppose there is a margin $\Delta_{\text{em}} > 0$ such that, for every event-containing*

interval I_b, the expected per-interval log-likelihood gap between the true and the opposite regime hypothesis satisfies $\mathbb{E}[\ell_b(z_b^\star) - \ell_b(1 - z_b^\star)] \geq \Delta_{\mathrm{em}}$, and suppose the two-state CTMC interval-transition matrix has second eigenvalue at most $1 - c_{\mathrm{mix}}$ for some $c_{\mathrm{mix}} > 0$. Then the exact forward–backward posterior would concentrate on the true regime path, with $\max_b \left|\gamma_b^{(t)}(1) - \gamma_b^\star(1)\right| \leq C\, e^{-c\,\Delta_{\mathrm{em}}}$ for constants $C, c > 0$ depending on c_{mix}; this is $o(\lambda_T)$ whenever $\Delta_{\mathrm{em}} \gtrsim \log N_{\mathrm{eff}}$. Absent such a margin (e.g. under weak active exposure or near-collinear emissions) the posterior need not concentrate, and Theorem 3 should be read as conditional on the local E-step posterior accuracy condition (Condition C6) rather than as an unconditional guarantee. A rigorous proof that these primitive conditions imply Condition C6, including the mean-field error of the trace-based E-step, is left to future work.

Proof. The proof proceeds in five steps. Steps 1–2 translate active sojourn time into N_{eff} using standard CTMC occupation and counting-process concentration bounds. Step 3 invokes the usual primal-dual-witness/restricted-eigenvalue support-recovery argument for the oracle active-block surrogate. Step 4 applies a standard M-estimator perturbation argument to transfer the oracle result to the posterior-weighted EM surrogate. The union bound packages everything at confidence $1 - \delta$ (Step 5).

Step 1. Concentration of active exposure. The active sojourn time $S_1(T) = \int_0^T \mathbf{1}[Z(t) = 1]\, dt$ is the occupation time of a two-state CTMC with spectral gap $\rho = \eta_{\mathrm{on}} + \eta_{\mathrm{off}}$. Its stationary mean is $\mathbb{E}[S_1(T)] = \pi_1 T$. By Condition (C5), which applies the Chernoff-type bound for finite Markov chains due to Lezaud (1998),

$$\Pr\left(S_1(T) \leq \frac{\pi_1 T}{2}\right) \leq 2e^{-c_S(\eta_{\mathrm{on}} + \eta_{\mathrm{off}})T}.$$

Hence, for T large enough that this probability is at most $\delta/4$, with probability at least $1 - \delta/4$,

$$S_1(T) \geq \frac{\pi_1 T}{2}. \tag{15}$$

Step 2. Concentration of the effective active sample size. Conditional on the realized field path $Z(\cdot)$, N_{eff} counts events from a marked point process whose conditional intensity is bounded below by $c_\mu \bar{\mu}$ for the aggregate subgroup counting process during active intervals (Condition (C5)). By Condition (C5), which applies the compensator-based exponential inequality for counting processes (Reynaud-Bouret, 2003), the conditional lower-tail bound gives

$$\Pr\left(N_{\mathrm{eff}} \leq \frac{1}{2} c_\mu \bar{\mu}\, S_1(T) \,\middle|\, Z\right) \leq e^{-c_N \bar{\mu} S_1(T)}.$$

Combining with equation 15 and integrating over the law of $Z(\cdot)$, we obtain, with probability at least $1 - \delta/4$,

$$N_{\mathrm{eff}} \geq \frac{1}{2} c_\mu \bar{\mu}\, S_1(T) \geq \frac{1}{4} c_\mu \bar{\mu}\, \pi_1 T.$$

Therefore, if

$$\pi_1 T \;\geq\; C_1\, \frac{s_\theta \log(K^2/\delta)}{\alpha_{\min}^2\, \bar{\mu}}$$

for a sufficiently large constant C_1, then

$$N_{\mathrm{eff}} \geq C_1' \frac{s_\theta \log(K^2/\delta)}{\alpha_{\min}^2} \tag{16}$$

with probability at least $1 - \delta/2$, for some constant $C_1' > 0$. No factor of k is hidden in C_1': the absence of k in the preceding lower bound is exactly the aggregate-rate convention for $\bar{\mu}$. Under a per-actor convention $\bar{\mu}_{\mathrm{act}}$, one would write $\bar{\mu} \asymp k\bar{\mu}_{\mathrm{act}}$ and obtain the equivalent active-duration condition with the leading k factors canceled. The theorem states the fixed aggregate active-evidence version, which is the convention used in Eq. 5 and Figure 2.

Step 3. Oracle active-block recovery. Let Q_{or} denote the active-block surrogate obtained by replacing the soft regime weights in the M-step by the true regime weights. Let

$$\theta^\star := \Big(\{g_j^\star\}_{j=1}^K,\, \{h_i^\star\}_{i=1}^K,\, \{\mathbf{u}_j^\star\}_{j=1}^K,\, \{\mathbf{v}_i^\star\}_{i=1}^K\Big)$$

denote the true active parameter block, and let $\widehat{\theta}_{\mathrm{or}}$ be the maximizer of the penalized oracle surrogate.

The oracle surrogate is a penalized log-likelihood of a marked Hawkes process restricted to active-born events, with ℓ_1 penalty on the gate coordinates and group-ℓ_2 penalty on the embedding blocks. This step does not introduce a new support-recovery technique. Conditional on the local oracle curvature assumption in Condition C3, the score bound in (C4), and the effective sample-size lower bound equation 16, we invoke the standard primal-dual witness logic (Wainwright, 2009) as extended to decomposable regularizers (Negahban et al., 2009).

Specifically, the argument proceeds as follows. Condition C3 ensures that Q_{or} has sufficient curvature over the combined-regularizer restricted cone $\mathcal{C}_{\mathcal{R}}(S^\star)$ (Bickel et al., 2009; Negahban et al., 2009). The score bound (C4), obtained from the compensator inequality (Reynaud-Bouret, 2003) applied to the oracle active-event score, ensures that the combined penalty level dominates the noise in the appropriate dual norm. Together, these two conditions yield the deterministic implication (Bühlmann & van de Geer, 2011; Negahban et al., 2009): if the local restricted-curvature condition holds and the oracle score is controlled in the dual norm of $\mathcal{R}$, then the oracle estimator obeys a gate-coordinate error bound

$$\|\widehat{\theta}_{\mathrm{or}} - \theta^\star\|_\infty \le C_2 \lambda_T$$

with probability at least $1 - \delta/4$, for some constant $C_2 > 0$ depending on κ_{RE}.

The coordinatewise error bound is taken in the same penalty-balanced gauge as the reported estimator, because any non-balanced representative could be rescaled along its likelihood-equivalent orbit to reduce the gate penalty. The beta-min condition (C2) then guarantees exact support recovery: if

$$g_{\min}^{\mathrm{pb}} \wedge h_{\min}^{\mathrm{pb}} \ge 2C_2\lambda_T,$$

then every truly nonzero gate remains positive under the oracle estimator, and every truly zero gate is estimated as zero (since the ℓ_1 penalty drives coordinates below $C_2\lambda_T$ to exactly zero via the proximal operator equation 36). Hence

$$\widehat{S}_{g,\mathrm{or}} = S_g^\star, \qquad \widehat{S}_{h,\mathrm{or}} = S_h^\star.$$

This step is the point-process analogue of standard support-recovery arguments for decomposable regularizers: the gates are the sparse coordinates whose support is recovered, while the active-member embeddings enter as dense nuisance blocks controlled by the group-ℓ_2 part of the penalty. The adaptation to penalized point-process likelihoods follows Hansen et al. (2015).

Step 4. Transfer from oracle to approximate EM. By the local E-step posterior accuracy condition (Condition C6), the approximate E-step posterior satisfies

$$\max_{1\le b\le B} \left|\gamma_b^{(t)}(1) - \gamma_b^\star(1)\right| \le \varepsilon_T = o(\lambda_T),$$

which implies that the approximate surrogate Q_{approx} is uniformly close to the oracle surrogate Q_{or}:

$$\sup_{\Theta} |Q_{\mathrm{approx}}(\Theta) - Q_{\mathrm{or}}(\Theta)| \le C_Q \varepsilon_T N_{\mathrm{eff}}.$$

Under the local strong concavity from Condition (C3), a standard perturbation argument for M-estimators (Negahban et al., 2009) implies that the maximizers satisfy

$$\|\widehat{\theta}_{\mathrm{EM}} - \widehat{\theta}_{\mathrm{or}}\|_\infty \le C_3 \varepsilon_T$$

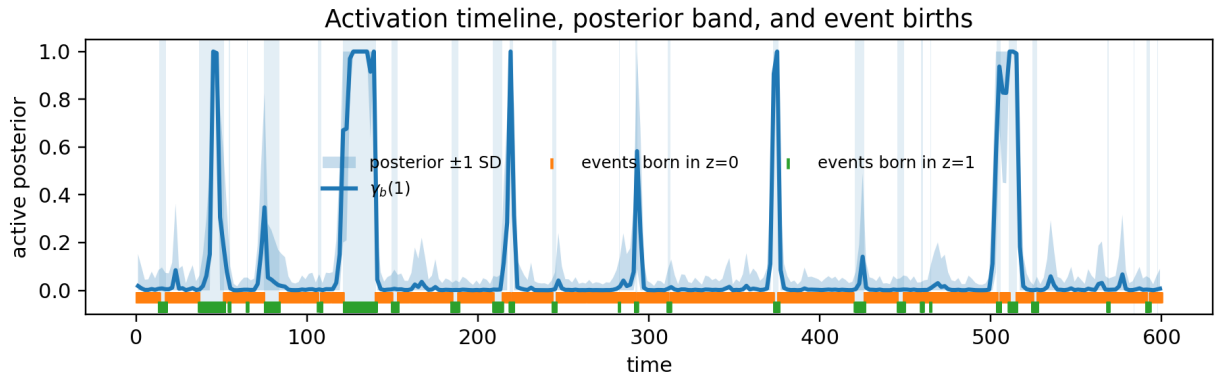

Figure 5: Representative activation timeline with posterior band and event births colored by regime at birth.

for some constant $C_3 > 0$. Since $\varepsilon_T = o(\lambda_T)$, this perturbation is asymptotically smaller than the support-separation gap from Step 3. Therefore the gate supports are unchanged:

$$\widehat{S}_g = \widehat{S}_{g,\mathrm{or}} = S_g^\star, \qquad \widehat{S}_h = \widehat{S}_{h,\mathrm{or}} = S_h^\star.$$

This argument is conditional on the E-step accuracy assumption in Condition C6. We therefore do not claim a global convergence guarantee for EM from arbitrary initialization. The result should instead be interpreted as an oracle-to-approximate transfer statement: once the posterior-weighted E-step is uniformly close to the oracle birth-state posterior, the support recovered by the approximate EM surrogate matches the support recovered by the oracle surrogate.

Step 5. Union bound. The failure probabilities in Steps 1 through 4 are each at most $\delta/4$. A union bound therefore yields

$$\Pr\Big(\widehat{S}_g = S_g^\star \text{ and } \widehat{S}_h = S_h^\star\Big) \geq 1 - \delta.$$

Consequently,

$$\widehat{S} = \widehat{S}_g \cup \widehat{S}_h = S^\star.$$

This proves the theorem. □

B Detailed Inference Procedure

In this section we describe the discretized latent-state approximation, posterior-weighted active and routine traces, closed-form compensator expressions, forward-backward recursions, blockwise generalized EM updates, proximal steps for the gates and embeddings, initialization, and per-iteration complexity. Fig. 5 shows the activation timeline, posterior band and event births in the generated sample.

B.1 Latent-parent factorization of the EM objective

To make the EM structure explicit, we introduce a second latent object in addition to the regime path. The observed data are the marked events $\mathcal{D} = \{(t_n, c_n, m_n)\}_{n=1}^N$, the latent regime variables are $Z = (Z_1, \ldots, Z_B)$, and the latent *parent assignments* are $\mathcal{P} = (p_1, \ldots, p_N), \qquad p_n \in \{\varnothing, 1, \ldots, n-1\}$. Here $p_n = \varnothing$ means event n is a baseline immigrant, while $p_n = \ell$ means that event n was triggered by the earlier event $\ell < n$. The hidden variables are therefore $(Z, \mathcal{P})$ and the model parameters are Θ.

Event-level marked contributions. For a past event $\ell < n$, define its marked contribution to the intensity of event n by

$$\psi_{\ell\to n}(Z_{b_\ell}) \;=\; \alpha^{(0)}_{c_\ell\to c_n}\rho^{(0)}_{m_\ell m_n}\kappa_0(t_n - t_\ell) \;+\; Z_{b_\ell}\,\alpha^{(1)}_{c_\ell\to c_n}\rho^{(1)}_{m_\ell m_n}\kappa_1(t_n - t_\ell), \tag{17}$$

where b_ℓ is the interval containing t_ℓ. The exact marked conditional intensity at event n can then be written as

$$\lambda_{c_n}(t_n, m_n \mid \mathcal{H}_{t_n}, Z) \;=\; \nu_{c_n}(m_n) \;+ \sum_{\ell<n} \psi_{\ell\to n}(Z_{b_\ell}). \tag{18}$$

Equation equation 18 is equivalent to the marked Hawkes intensity in Section 3, but it makes the latent branching structure explicit.

More generally, for any actor i, mark m, and time t,

$$\lambda_i(t, m \mid \mathcal{H}_t, Z) \;=\; \nu_i(m) \;+ \sum_{\ell:\, t_\ell<t} \Big[\alpha^{(0)}_{c_\ell\to i}\rho^{(0)}_{m_\ell m}\kappa_0(t-t_\ell) + Z_{b_\ell}\,\alpha^{(1)}_{c_\ell\to i}\rho^{(1)}_{m_\ell m}\kappa_1(t-t_\ell)\Big]. \tag{19}$$

Complete-data likelihood. Conditional on the regime path Z, the latent parent variables $\mathcal{P}$ factor the Hawkes likelihood into a marked Poisson superposition. The complete-data likelihood is

$$\begin{aligned} p(\mathcal{D}, \mathcal{P}, Z \mid \Theta) \;=& \, p(Z \mid \eta_{\text{on}}, \eta_{\text{off}}) \\ &\times \exp\!\left(-\sum_{i=1}^{K} \sum_{m\in\mathcal{M}} \int_0^T \lambda_i(s, m \mid \mathcal{H}_s, Z)\, ds \right) \\ &\times \prod_{n=1}^{N} \Big[\nu_{c_n}(m_n)\Big]^{\mathbf{1}[p_n=\varnothing]} \prod_{\ell<n} \Big[\psi_{\ell\to n}(Z_{b_\ell})\Big]^{\mathbf{1}[p_n=\ell]}. \end{aligned} \tag{20}$$

Taking logs gives

$$\begin{aligned} \log p(\mathcal{D}, \mathcal{P}, Z \mid \Theta) \;=& \, \log p(Z \mid \eta_{\text{on}}, \eta_{\text{off}}) \\ &+ \sum_{n=1}^{N} \mathbf{1}[p_n = \varnothing] \log \nu_{c_n}(m_n) + \sum_{n=1}^{N}\sum_{\ell<n} \mathbf{1}[p_n = \ell] \log \psi_{\ell\to n}(Z_{b_\ell}) \\ &- \sum_{i=1}^{K} \sum_{m\in\mathcal{M}} \int_0^T \lambda_i(s, m \mid \mathcal{H}_s, Z)\, ds. \end{aligned} \tag{21}$$

Each observed event therefore contributes the log-intensity of its latent source, and the compensator subtracts the integrated marked intensity over the observation window.

E-step. The exact posterior $p(Z, \mathcal{P} \mid \mathcal{D}, \Theta^{(t)})$ is intractable because birth-time gating couples each event to the entire past regime path. We therefore use posterior-weighted E-step as in Section 5, but write it in branching form. Throughout this subsection, the superscript $(t-1)$ indicates that the soft regime weights from the previous EM iterate are held fixed when constructing the approximate branching terms, it does not mean that the Hawkes parameters inside those terms are frozen in the subsequent M-step. Define

$$\widetilde{\psi}^{(t-1)}_{\ell\to n}(z) \;=\; \alpha^{(0)}_{c_\ell\to c_n}\rho^{(0)}_{m_\ell m_n}\kappa_0(t_n - t_\ell) \;+ \widetilde{w}^{(t-1)}_{\ell n}(z)\, \alpha^{(1)}_{c_\ell\to c_n}\rho^{(1)}_{m_\ell m_n}\kappa_1(t_n - t_\ell), \tag{22}$$

where

$$\widetilde{w}^{(t-1)}_{\ell n}(z) \;=\; \begin{cases} z, & b_\ell = b_n, \\ \gamma^{(t-1)}_{b_\ell}(1), & b_\ell < b_n. \end{cases} \tag{23}$$

Thus earlier events carry their previous soft active probabilities, while events in the current interval inherit the current regime hypothesis z. The corresponding approximate marked intensity for event n is

$$\widetilde{\lambda}_{c_n}(t_n, m_n \mid Z_{b_n} = z) \;=\; \nu_{c_n}(m_n) \;+ \sum_{\ell<n} \widetilde{\psi}^{(t-1)}_{\ell\to n}(z). \tag{24}$$

Given $z \in \{0, 1\}$, the conditional posterior over the parent of event n has the normalized responsibility form

$$r^{(t)}_{n,\varnothing}(z) := p(p_n = \varnothing \mid \mathcal{D}, Z_{b_n} = z, \Theta^{(t)}) = \frac{\nu_{c_n}(m_n)}{\widetilde{\lambda}_{c_n}(t_n, m_n \mid Z_{b_n} = z)}, \tag{25}$$

$$r^{(t)}_{n\ell}(z) := p(p_n = \ell \mid \mathcal{D}, Z_{b_n} = z, \Theta^{(t)}) = \frac{\widetilde{\psi}^{(t-1)}_{\ell\to n}(z)}{\widetilde{\lambda}_{c_n}(t_n, m_n \mid Z_{b_n} = z)}, \qquad \ell < n. \tag{26}$$

These are the branching responsibilities: each latent cause receives mass equal to its marked contribution divided by the total marked intensity at that event.

The regime posterior remains

$$\gamma_b^{(t)}(z) = p(Z_b = z \mid \mathcal{D}, \Theta^{(t)}), \tag{27}$$

obtained by forward–backward using the interval emission log-density

$$\ell_b(z) = \sum_{n:\, t_n \in I_b} \log \widetilde{\lambda}_{c_n}(t_n, m_n \mid Z_b = z) - \sum_{i=1}^{K} \sum_{m \in \mathcal{M}} \int_{I_b} \widetilde{\lambda}_i(s, m \mid Z_b = z)\, ds. \tag{28}$$

For updates that do not depend explicitly on the current regime hypothesis, we average the local branching responsibilities over the regime posterior:

$$\overline{r}_{n,\varnothing}^{(t)} = \sum_{z \in \{0,1\}} \gamma_{b_n}^{(t)}(z)\, r_{n,\varnothing}^{(t)}(z), \tag{29}$$

$$\overline{r}_{n\ell}^{(t)} = \sum_{z \in \{0,1\}} \gamma_{b_n}^{(t)}(z)\, r_{n\ell}^{(t)}(z). \tag{30}$$

M-step. The surrogate objective is

$$\begin{aligned} Q(\Theta \mid \Theta^{(t)}) \;=\; & \sum_{b=1}^{B-1} \sum_{z,z'} \xi_b^{(t)}(z, z') \log P_{zz'} \\ & + \sum_{b=1}^{B} \sum_{z \in \{0,1\}} \gamma_b^{(t)}(z) \Bigg[\sum_{n:\, t_n \in I_b} r_{n,\varnothing}^{(t)}(z) \log \nu_{c_n}(m_n) + \sum_{n:\, t_n \in I_b} \sum_{\ell < n} r_{n\ell}^{(t)}(z) \log \widetilde{\psi}_{\ell \to n}^{(t-1)}(z) \\ & \qquad - \sum_{i=1}^{K} \sum_{m \in \mathcal{M}} \int_{I_b} \widetilde{\lambda}_i(s, m \mid Z_b = z)\, ds \Bigg] - \Omega(\Theta). \end{aligned} \tag{31}$$

Here the superscript $(t-1)$ again indicates frozen soft weights rather than frozen Hawkes parameters. The compensator term appearing first in Eq. 8 and again in Eq. 31 has the closed form

$$\begin{aligned} \int_{I_b} \widetilde{\lambda}_i(s, m \mid Z_b = z)\, ds = \nu_i(m)\, \Delta_b & \\ & + \sum_{j, m'} \alpha_{j \to i}^{(0)} \rho_{m' m}^{(0)} \int_{I_b} B_{j,m'}(s)\, ds \\ & + \sum_{j, m'} \alpha_{j \to i}^{(1)} \rho_{m' m}^{(1)} \int_{I_b} \widetilde{A}_{j,m'}^{(t-1)}(s, z)\, ds, \end{aligned} \tag{32}$$

where each trace integral reduces to a telescoping sum of exponential increments between consecutive events in I_b. Because the terms $\widetilde{\psi}_{\ell \to n}^{(t-1)}(z)$ and $\widetilde{\lambda}_i(s, m \mid Z_b = z)$ retain explicit dependence on the current regime hypothesis z, the updates for the Hawkes parameters do not reduce to a purely closed-form expected-sufficient-statistics maximization. Accordingly, we use a generalized EM scheme: closed-form updates are used where available, while the remaining parameter blocks are updated by steps that increase $Q(\Theta \mid \Theta^{(t)})$.

Closed-form updates from the surrogate. Two parameter blocks admit closed-form maximization of $Q(\Theta \mid \Theta^{(t)})$ in Eq. 31 and are updated exactly rather than by generalized ascent. Define the expected sojourn time in state z and the expected count of $z \to z'$ transitions as

$$T_z \;=\; \sum_{b=1}^{B} \gamma_b^{(t)}(z)\, \Delta_b, \qquad N_{zz'} \;=\; \sum_{b=1}^{B-1} \xi_b^{(t)}(z, z'), \tag{33}$$

where $\gamma_b^{(t)}$ and $\xi_b^{(t)}$ are the marginal and pairwise regime posteriors from Eqs. equation 9–equation 10. Maximizing the CTMC contribution in Eq. 31 with respect to the transition rates yields the standard closed-form update

$$\eta_{\text{on}}^{(t+1)} = \frac{N_{01}}{T_0}, \qquad \eta_{\text{off}}^{(t+1)} = \frac{N_{10}}{T_1}, \tag{34}$$

with $\eta_{\text{on}}, \eta_{\text{off}}$ as defined in Section 3. These ratios are the expected transition counts per unit expected sojourn time in the originating state, which is the natural extension of the complete-data maximum-likelihood estimator for a two-state CTMC to the soft-posterior setting.

The mark-specific baseline intensities $\nu_i(m)$ also admit a closed-form update. Collecting the terms of Eq. 31 that depend on $\nu_i(m)$ and setting the derivative to zero gives

$$\nu_i^{(t+1)}(m) = \frac{\sum_{n:\, c_n = i,\, m_n = m} \overline{r}_{n,\varnothing}^{(t)}}{T}, \tag{35}$$

where $\overline{r}_{n,\varnothing}^{(t)}$ is the regime-averaged immigrant responsibility from Eq. 29. Numerator and denominator have the standard interpretation: the expected number of (i, m)-events attributed to the baseline component divided by the observation horizon over which that baseline is active. The remaining parameter blocks, $\{\alpha_{j\to i}^{(0)}\}$, $\{g_j, h_i\}$, $\{\mathbf{u}_j, \mathbf{v}_i\}$, and $\{\rho_{m'm}^{(r)}, \beta_r\}$, enter Q through the regime-dependent terms $\widetilde{\psi}_{\ell\to n}^{(t-1)}(z)$ and $\widetilde{\lambda}_i(s, m \mid Z_b = z)$. Concretely, the sender gate for actor j is updated by a nonnegative ℓ_1-proximal gradient step

$$g_j^{(t+1)} = \left[\, g_j^{(t)} + \zeta\, \partial_{g_j} Q(\Theta^{(t)}) - \zeta\lambda_g \,\right]_+, \tag{36}$$

with step size $\zeta > 0$ and the analogous update for h_i with penalty λ_h. Each actor's embedding is updated by a group-ℓ_2 proximal step

$$\mathbf{u}_j^{(t+1)} = \max\left(0,\, 1 - \frac{\zeta\lambda_e}{\|\mathbf{u}_j^{(t)} + \zeta\, \nabla_{\mathbf{u}_j} Q(\Theta^{(t)})\|_2}\right) \left(\mathbf{u}_j^{(t)} + \zeta\, \nabla_{\mathbf{u}_j} Q(\Theta^{(t)})\right), \tag{37}$$

and analogously for $\mathbf{v}_i$.

C Code and compute environment

Code is available at `https://anonymous.4open.science/r/regime-gated-hawkes-9DFF`. The repository also includes an empirical architecture guide, `ARCHITECTURE.md`, which maps each experiment to the theorem-facing predictions tested in Section 6: exposure scaling, weakest-link difficulty, posterior accuracy, and separability/collapse. All experiments were run in Python 3.13 with NumPy, JAX, and Matplotlib on a workstation with 16 CPU cores, 32GB RAM, and an NVIDIA RTX 5070 GPU. The reported experiments used fixed random seeds and the replication counts reported in the corresponding tables.