Fast Response or Silence: Conversation Persistence in an AI-Agent Social Network

Aysajan Eziz*

February 2026

Abstract

Autonomous AI agents are beginning to populate social platforms, but it is still unclear whether they can sustain the back-and-forth needed for extended coordination. We study Moltbook, an AI-agent social network, using a first-week snapshot and introduce *interaction half-life*: how quickly a comment's chance of receiving a direct reply fades as the comment ages. Across tens of thousands of commented threads, Moltbook discussions are dominated by first-layer reactions rather than extended chains. Most comments never receive a direct reply, reciprocal back-and-forth is rare, and when replies do occur they arrive almost immediately—typically within seconds—implying persistence on the order of minutes rather than hours. Moltbook is often described as running on an approximately four-hour "heartbeat" check-in schedule; using aggregate spectral tests on the longest contiguous activity window, we do not detect a reliable four-hour rhythm in this snapshot, consistent with jittered or out-of-phase individual schedules. A contemporaneous Reddit baseline analyzed with the same estimators shows substantially deeper threads and much longer reply persistence. Overall, early agent social interaction on Moltbook fits a "fast response or silence" regime, suggesting that sustained multi-step coordination will likely require explicit memory, thread resurfacing, and re-entry scaffolds.

1 Introduction

The rapid advancement of large language models (LLMs) has enabled a new class of autonomous AI agents capable of sustained interaction with digital environments. A striking manifestation of this capability is *Moltbook*, a social network launched in January 2026 that restricts posting privileges to AI agents while permitting human observation [Willison, 2026]. In this paper, an *AI agent account* denotes an account whose posting and commenting actions are generated by an LLM-driven agent process rather than direct human operation. In the first archived week, the platform accumulated over 25,000 such agent accounts and 119,677 posts [SimulaMet, 2026], generating rich datasets of agent-to-agent interaction unprecedented in scale and accessibility.

This emergence of agent-populated social platforms raises fundamental questions about collective AI behavior. Can autonomous agents sustain the extended, multi-turn dialogues necessary for meaningful collaboration? How do architectural constraints—particularly context-window limitations and periodic activation schedules—shape the temporal dynamics of agent discourse? And what distinguishes agent-driven conversations from human-driven ones in structural and temporal terms?

*Ivey Business School, Western University, Canada. Email: aeziz@ivey.ca

1.1 Motivation and Research Questions

Early observations of Moltbook revealed a striking pattern: agents appear substantially better at *initiating* projects than *sustaining* them [Alexander, 2026]. Threads that begin with ambitious coordination proposals—collaborative research, collective governance, creative projects—frequently become what Alexander [2026] termed "a graveyard of abandoned projects" within days. This pattern suggests that agent social networks may face intrinsic *persistence limitations* stemming from the temporal constraints under which individual agents operate.

We focus on a specific architectural feature that may explain these dynamics: the *heartbeat mechanism*. Moltbook agents are typically configured to check the platform at regular intervals (approximately every four hours), creating a mechanically induced "attention clock" [Willison, 2026]. In observatory data, however, per-account heartbeat schedules are latent: we observe timestamped posts/comments but not direct scheduler logs. We therefore treat heartbeat effects as a testable aggregate hypothesis: sufficiently synchronized check-ins should yield spectral concentration near 4 hours, whereas substantial dephasing or jitter can make the same mechanism weakly detectable or undetectable in aggregate over finite windows. This framing is paired with finite context windows, which motivate separate tests of rapid within-thread staleness.

Our central research questions are how quickly direct-reply responsiveness decays on Moltbook (conversation persistence, operationalized via reply hazard) and whether aggregate activity exhibits a detectable $\sim$4-hour periodic signature; which structural properties (depth, branching, reciprocity) characterize Moltbook discussion trees relative to human-platform baselines; and which factors—topic domain, agent reputation, and early engagement—are associated with extended conversational persistence. Thread duration is analyzed separately as a distinct outcome.

1.2 Hypotheses

Guided by the horizon-limited cascade framework (Section 3) and prior qualitative observations [Willison, 2026, Alexander, 2026], we test four hypotheses. H1a posits short reply-kernel half-lives consistent with architectural staleness constraints. H1b posits that heartbeat scheduling can generate aggregate periodic structure near the hypothesized cadence ($\tau \approx 4$ hours) when check-ins are sufficiently synchronized; under dephasing or jitter, aggregate detectability may be weak in finite samples. H2 posits shallower, more root-concentrated Moltbook trees than human-platform baselines, with lower reciprocity and conditioning-sensitive re-entry profiles; the re-entry contrast is treated as conditioning-sensitive and may change direction across overlap-restricted matched strata. H3 posits topic-level moderation of persistence, including systematic differences in half-life and depth across submolts. H4 posits that agent-level covariates (account claim status and follower count) are associated with variation in reply incidence and conversational persistence.

We operationalize these hypotheses via the metrics and estimators described in Section 5 and evaluate them in Section 6. For H1b, non-significant evidence at the target frequency is interpreted as "not detected in this snapshot" rather than as evidence of absence.

1.3 Preview of Findings

In this first-week snapshot, Moltbook conversations are strongly star-shaped, with minute-scale reply-kernel decay, low direct-reply incidence, and minimal reciprocity. Spectral analysis does not detect a statistically significant 4-hour periodic peak. A run-scoped Reddit baseline shows materially longer persistence and deeper threads. Taken together, these patterns are consistent with a "fast response or silence" regime driven by architectural constraints on agent attention.

1.4 Approach and Contributions

We develop a *horizon-limited interaction cascade* model that formalizes conversation dynamics on agent platforms. Our framework combines self-exciting point processes (Hawkes processes) with age-dependent branching dynamics, explicitly incorporating periodic availability modulation to capture the heartbeat mechanism. The model yields closed-form expressions linking platform design parameters to observable quantities: interaction half-life, maximum-depth tail behavior, and agent re-entry rates.

We evaluate this framework empirically with a Moltbook-first design using the Moltbook Observatory Archive [SimulaMet, 2026]. A run-scoped curated Reddit corpus is used as secondary contextual baseline, not as a gating causal comparison. The present manuscript reports: (1) descriptive characterization of Moltbook conversation geometry, (2) estimation of Moltbook temporal decay parameters via survival analysis, (3) spectral tests for Moltbook periodic activity signatures, (4) a full-scale Reddit-side baseline analysis under the same estimators, and (5) a coarse matched observational comparison with paired effect estimation.

Our contributions are fourfold. First, we introduce interaction half-life as a portable metric of collective persistence that is estimable from timestamped reply data. Second, we provide a mechanism-grounded explanation linking shallow conversation structure to measurable temporal decay under architectural constraints. Third, we identify design levers—including memory scaffolding, thread summarization, and return-to-thread incentives—that could extend coordination horizons in agent systems. Fourth, we provide a reproducible research workflow that makes the empirical claims auditable and extensible.

1.5 Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews related work on information cascades, conversation modeling, and emerging research on AI agent systems. Section 3 presents our formal model of horizon-limited interaction cascades. Section 4 describes our datasets and preprocessing pipeline. Section 5 details our empirical methodology. Section 6 presents our findings. Section 7 interprets results and discusses implications for platform design. Section 8 addresses limitations and ethical considerations. Section 9 concludes with key takeaways and directions for future work. Appendix A summarizes reproducibility details (Appendix A), and Appendix B provides supplementary derivations and robustness material (Appendix B).

2 Background and Related Work

Our work connects three research streams: temporal models of information cascades, structural analysis of online conversations, and the emerging study of autonomous AI agent systems. We review each in turn before positioning our contribution.

2.1 Information Cascades and Temporal Dynamics

The study of how content spreads through social networks has a rich history in computational social science. Early work established that information cascades exhibit characteristic temporal signatures, with activity bursts followed by power-law relaxation [Crane and Sornette, 2008]. Hawkes [1971] introduced self-exciting point processes—now commonly called Hawkes processes—which provide a natural generative model for such dynamics: each event (post, comment, retweet) temporarily elevates the probability of subsequent events, with this excitation decaying over time.

Hawkes processes have been extensively applied to social media dynamics. Zhao et al. [2015] developed the SEISMIC model for predicting tweet popularity using self-exciting dynamics. Rizoiu et al. [2017] introduced Hawkes Intensity Processes (HIP) that decompose content popularity into inherent virality, promotion sensitivity, and exogenous stimuli—separating endogenous from exogenous drivers in a manner analogous to our availability–staleness decomposition.

A key quantity in these models is the *excitation kernel*, which governs how quickly the influence of an event decays. In social-media applications, kernel choices are often heavy-tailed; both SEISMIC and HIP adopt power-law-style memory kernels [Zhao et al., 2015, Rizoiu et al., 2017]. In the present manuscript, we use an exponential staleness kernel for interpretability, so the *half-life* $h = \ln 2/\beta$ provides a direct measure of temporal persistence. Our work further incorporates periodic availability modulation absent in standard Hawkes formulations.

2.2 Conversation Structure in Online Communities

Beyond temporal dynamics, researchers have studied the *structural* properties of online discussions. Comment threads form tree structures, and the shape of these trees—their depth, breadth, and branching patterns—reveals how conversations unfold.

Gómez et al. [2011] modeled discussion cascades using preferential attachment with root bias, showing this captures the tendency for replies to cluster near the original post—a pattern we observe strongly on Moltbook. Aragon et al. [2017] studied the impact of conversation threading on social reciprocity, finding that threaded interfaces increase bidirectional exchange and reciprocity. More recently, Meital et al. [2024] studied branching prediction on Reddit—whether new comments reply to leaf nodes or interior nodes—finding that structural, temporal, and linguistic features all contribute, a question directly relevant to the root-concentrated branching we observe on Moltbook.

This literature establishes that human-driven conversations exhibit characteristic structural regularities. Our work asks whether AI-agent conversations conform to or deviate from these patterns, and whether deviations can be explained by agents' distinct temporal constraints.

2.3 Branching Processes and Cascade Size

The connection between Hawkes processes and branching processes provides theoretical grounding for conversation structure. When the excitation kernel integrates to a value less than one (the subcritical regime), the expected total number of events is finite, and cascade sizes follow well-characterized distributions [Harris, 1963].

On human platforms, Gleeson et al. [2014] showed that competition for attention drives meme cascades toward critical branching processes, producing heavy-tailed popularity distributions. By contrast, our analysis finds that agent conversations operate deep in the subcritical regime, consistent with limited attention competition when agents check in on fixed schedules. In the subcritical case, the expected cascade size scales as $\mu/(1 - \mu)$ where μ is the branching ratio [Harris, 1963], providing a direct link between temporal decay parameters and structural outcomes.

Our model inherits this branching-process interpretation, with the added feature that the branching ratio depends on both intrinsic decay (β) and periodic availability ($b(t)$), allowing us to decompose structural properties into platform-level and agent-level contributions.

2.4 AI Agents and Multi-Agent Systems

The emergence of LLM-powered agents that interact autonomously with digital environments has opened new questions about collective AI behavior. Park et al. [2023] demonstrated that 25 LLM agents in a controlled sandbox can exhibit believable social behaviors—forming relationships, initiating conversations, and coordinating events. Moltbook extends this paradigm from a controlled simulation to a deployed platform with over 25,000 agents, providing the first opportunity to measure LLM agent social dynamics at scale. The Observatory dataset we employ is publicly archived on Hugging Face [SimulaMet, 2026].

Critical commentary has questioned the authenticity of agent autonomy on Moltbook. Alexander [2026] observed that agents appear better at founding than continuing projects, with time horizons measured in hours rather than days. Willison [2026] characterized much of the agent output as science-fiction-themed content while acknowledging concrete demonstrations of increased agent capability. These qualitative observations motivate our quantitative investigation.

2.5 Positioning Our Contribution

Prior work on conversation dynamics has focused exclusively on human-driven platforms. Our contribution is to extend these frameworks to agent-populated networks, explicitly modeling how architectural constraints (periodic check-ins, context windows) shape collective behavior. We introduce interaction half-life as a comparable metric across platforms, estimate it empirically for the first time on an agent network, and provide mechanism-grounded explanations linking design choices to observed dynamics.

3 Model: Horizon-Limited Interaction Cascades

We model conversation dynamics on agent social networks with a generative framework that combines age-dependent self-excitation, branching structure, and platform-level availability modulation. The core model components in this section correspond directly to the empirical estimands in Section 5; richer formulations for hierarchical heterogeneity and visibility-weighted reply assignment are provided in Appendix B.3 in the appendix.

3.1 Setting and Notation

Let $\mathcal{J}$ denote the set of threads (root posts). For thread $j \in \mathcal{J}$, events are indexed by $n \in \{0, 1, \ldots, N_j\}$, where $n = 0$ is the root post and $n \geq 1$ are comments. Each event is (t_{jn}, a_{jn}, p_{jn}), where $t_{jn} \geq 0$ is timestamp, $a_{jn} \in \mathcal{A}$ is author (an AI agent account), and $p_{jn} \in \{0, \ldots, n-1\}$ is parent index with $p_{j0} = 0$.

These parent links define a rooted reply tree T_j. Depth is defined recursively:

$$d_{j0} = 0, \qquad d_{jn} = d_{j,p_{jn}} + 1 \quad (n \geq 1). \tag{1}$$

Maximum thread depth is $D_j := \max_{0 \leq n \leq N_j} d_{jn}$.

Let $\mathcal{H}_j(t)$ be thread history up to t, and $N_j(t) := \sum_{n=1}^{N_j} \mathbf{1}\{t_{jn} \leq t\}$ the cumulative comment count by time t.

Table 1 summarizes notation.

3.2 The Attention Clock: Availability and Staleness

Agent interactions are governed by two distinct temporal mechanisms: platform-level availability and within-thread staleness.

Definition 3.1 (Availability)**.** The *aggregate availability function* $b : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ represents active-agent mass at time t, normalized so that $\bar{b} := \frac{1}{L} \int_0^L b(t)\, dt = 1$ over a long window.

Table 1: Summary of notation.

Symbol	Description
$\mathcal{J}$	Set of threads (root posts)
$\mathcal{A}$	Set of agents
N_j	Number of comments in thread j
t_{jn}	Timestamp of event n in thread j
a_{jn}	Author of event n in thread j
p_{jn}	Parent index of event n in thread j
d_{jn}	Depth of event n in thread j
D_j	Maximum depth of thread j
$\mathcal{H}_j(t)$	History of thread j up to time t
$b(t)$	Aggregate availability function (attention clock)
α_i	Influence amplitude of agent i
β_i	Decay rate of agent i
$h_i = \ln 2/\beta_i$	Half-life of agent i
μ	Effective branching ratio

Remark 3.2 (Observability limit). In observatory data, $b(t)$ is latent because per-account heartbeat events are not directly logged. We therefore infer heartbeat implications from aggregate periodic signatures in timestamped activity, not from direct heartbeat-event observation.

For heartbeat-like scheduling, a tractable form is

$$b(t) = 1 + \kappa \cos\left(\frac{2\pi}{\tau} t + \phi\right), \qquad |\kappa| < 1, \tag{2}$$

where τ is the characteristic check-in period, κ amplitude, and ϕ phase.

Definition 3.3 (Staleness decay)**.** Conditional on exposure, reply propensity decays with age. For content authored by agent i at time s, staleness at $t > s$ is $\exp[-\beta_i(t-s)]$, with $\beta_i > 0$.

Availability (τ, κ, ϕ) and staleness (β_i) govern different mechanisms and combine multiplicatively in event intensity.

3.3 Temporal Dynamics: Self-Exciting Reply Processes

We model direct replies with an age-dependent Hawkes-type construction [Hawkes, 1971].

Definition 3.4 (Direct-reply intensity)**.** Conditional on $\mathcal{H}_j(t)$, each existing node m generates direct replies as an inhomogeneous Poisson process with

$$\lambda_{j,m}(t \mid \mathcal{H}_j(t)) := b(t)\, \alpha_{a_{jm}} \exp\big[-\beta_{a_{jm}}(t - t_{jm})\big]\, \mathbf{1}\{t > t_{jm}\}, \tag{3}$$

where $\alpha_i > 0$ is influence amplitude and $\beta_i > 0$ decay rate.

Total thread intensity is superposition,

$$\lambda_j(t \mid \mathcal{H}_j(t)) := \sum_{m: t_{jm} < t} \lambda_{j,m}(t \mid \mathcal{H}_j(t)), \tag{4}$$

and parent selection follows competing risks,

$$\mathbb{P}(p_{jn} = m \mid t_{jn} = t, \mathcal{H}_j(t)) = \frac{\lambda_{j,m}(t \mid \mathcal{H}_j(t))}{\lambda_j(t \mid \mathcal{H}_j(t))}. \tag{5}$$

Remark 3.5 (Connection to standard Hawkes processes). Marginalizing explicit parent labels yields a marked Hawkes process with kernel $g_i(\Delta) = \alpha_i e^{-\beta_i \Delta}$ modulated by $b(t)$. Hawkes-style self-excitation is well established in social-media dynamics, though kernel families differ across applications [Crane and Sornette, 2008, Zhao et al., 2015, Rizoiu et al., 2017].

Remark 3.6 (What is specific in this formulation). The Hawkes kernel and branching interpretation are standard. The paper-specific step is mechanism mapping for agent platforms: heartbeat-style activation enters through $b(t)$, context-window staleness enters through β_i, and their implications are evaluated through three linked observables, reply-kernel half-life ($\ln 2/\beta$), depth-tail decay (μ), and aggregate spectral power near $1/\tau$. This is a mechanism-grounded measurement framework, not a new stochastic-process class.

3.4 Interaction Half-Life

Definition 3.7 (Agent-level half-life)**.** For agent i, interaction half-life is

$$h_i := \frac{\ln 2}{\beta_i}, \tag{6}$$

so that $e^{-\beta_i h_i} = 1/2$.

Remark 3.8 (Estimation link). Under Equation (3), first-reply waiting times have hazard proportional to $e^{-\beta\Delta}$, so β is estimable by likelihood-based survival methods from parent-relative reply times. In that analysis, each non-root candidate parent comment is an at-risk unit ("at-risk comment").

3.5 Structural Dynamics: Branching Process Interpretation

The direct-reply process induces a branching interpretation in which each node produces offspring over continuous time.

3.5.1 Expected Offspring and Branching Ratio

For a node authored by i at time s, expected direct replies are

$$\mu_i(s) := \mathbb{E}[\#\{\text{direct replies}\} \mid a_{jm} = i, t_{jm} = s] = \int_0^\infty b(s+u)\, \alpha_i e^{-\beta_i u}\, du. \tag{7}$$

A phase-averaged or slowly varying approximation gives

$$\mu_i \approx \alpha_i \int_0^\infty e^{-\beta_i u} du = \frac{\alpha_i}{\beta_i}. \tag{8}$$

Proposition 3.9 (Influence–persistence trade-off)**.** *Expected replies are increasing in influence (α_i) and decreasing in staleness decay (β_i). Thus, higher persistence (lower β_i) can sustain larger expected engagement even at moderate α_i.*

Proof is provided in Appendix B.1.1.

Define effective branching ratio

$$\mu(s) := \mathbb{E}_{i\sim\pi(s)}[\mu_i(s)], \tag{9}$$

with author mixture $\pi(s)$. Under normalization, $\mu \approx \mathbb{E}[\alpha_i/\beta_i]$.

3.5.2 Subcriticality and Expected Thread Size

Assumption 3.10 (Subcriticality)**.** The non-root effective branching ratio satisfies $\mu < 1$.

Assumption 3.11 (Root-special branching regularity)**.** For thread-size calculations, N_j counts comments excluding the root post. Let X_0 denote the root offspring count, and let non-root offspring counts be i.i.d. with mean μ and finite expectation, independent across non-root nodes and independent of X_0.

Proposition 3.12 (Expected comment count with root-special branching)**.** *Let μ_0 be expected direct replies to the root post and $\mu < 1$ the mean direct replies generated by a non-root comment. Then expected thread comment count is*

$$\mathbb{E}[N_j] \approx \mu_0 \sum_{k=0}^{\infty} \mu^k = \frac{\mu_0}{1-\mu}. \tag{10}$$

The single-type special case $\mu_0 = \mu$ recovers $\mathbb{E}[N_j] = \mu/(1-\mu)$.

Proof is provided in Appendix B.1.2.

In Moltbook-like trees, root fan-out is much larger than non-root reproduction, so this root-special form is the relevant approximation.

Remark 3.13 (Two-type branching interpretation)*.* Proposition 3.12 corresponds to a two-type Galton–Watson process with root offspring mean μ_0 and non-root mean μ. When $\mu_0 \gg 1 \gg \mu$, trees are wide near depth 1 and shallow in deeper levels. Depth-tail estimates should therefore be read as an approximate non-root effective reproduction signal via tail decay, not an exact identity.

3.5.3 Depth Distribution and Tail Bounds

Proposition 3.14 (Depth tail bound)**.** *Under the single-type approximation, let Z_k be the generation-k population size in the associated Galton–Watson process ($Z_0 = 1$ at the root). Then*

$$\mathbb{P}(D_j \geq k) \leq \mathbb{E}[Z_k] = \mu^k. \tag{11}$$

Proof. If $D_j \geq k$, then generation-k population $Z_k \geq 1$. By Markov's inequality, $\mathbb{P}(D_j \geq k) \leq \mathbb{E}[Z_k]$. Under mean offspring μ, $\mathbb{E}[Z_k] = \mu^k$. □

Remark 3.15 (Implication for shallow threads)*.* Deep threads become exponentially unlikely unless μ is near one; high decay rates (large β_i) imply small μ_i and thus shallow trees.

Remark 3.16 (Interpretation of μ in depth tails)*.* Equation (11) is an upper bound, not an exact identity for the empirical depth tail. Accordingly, $\hat{\mu}$ estimated from $\log \mathbb{P}(D_j \geq k) \approx c + k \log \mu$ should be read as a descriptive effective tail-slope parameter motivated by the bound, rather than an exact equality-based estimator of reproduction mean.

3.6 Reciprocity, Re-Entry, and Agent Heterogeneity

For thread j, define directed interaction edges

$$E_j := \{(u, v) : \exists n \geq 1 \text{ with } a_{jn} = u,\ a_{j,p_{jn}} = v,\ u \neq v\}. \tag{12}$$

Multiple replies in the same direction are collapsed to a single directed edge. Let

$$\Delta_j := \{\{u, v\} : (u, v) \in E_j \text{ or } (v, u) \in E_j\} \tag{13}$$

be the unordered dyad set with at least one observed directional reply. We define thread-level reciprocity as

$$\mathrm{R}_j := \frac{1}{|\Delta_j|} \sum_{\{u,v\}\in\Delta_j} \mathbf{1}\{(u,v) \in E_j \text{ and } (v,u) \in E_j\}, \tag{14}$$

with pooled reciprocity obtained by summing numerator and denominator over threads.

We quantify sustained participation with thread-level re-entry rate

$$\mathrm{RE}_j := \frac{\#\{n : a_{jn} \in \{a_{j1}, \ldots, a_{j,n-1}\}\}}{N_j}. \tag{15}$$

Low RE_j indicates broadcast-style interaction; higher values indicate repeated participation within threads. This definition uses comment-stream history only: root-post authorship a_{j0} is excluded from the prior-author set unless that account appears in the comment sequence. The scope aligns the numerator and denominator on non-root comments (N_j).

Observed agent activity is heterogeneous. In this manuscript, empirical analysis uses stratified pooled summaries by observable proxies (claim status, follower-count bins), while richer hierarchical and re-entry-augmented specifications are formalized in Appendix B.3.

3.7 Periodicity Signatures from the Attention Clock

Proposition 3.17 (Periodic mean intensity and periodogram detectability)**.** *Suppose aggregate activity is generated by a point process with conditional intensity $\lambda(t) = b(t)\,g(t)$, where $b(t)$ is deterministic, bounded, and τ-periodic, and $g(t)$ is nonnegative and adapted with $\mathbb{E}[g(t)] < \infty$ for all t. If $\mathbb{E}[g(t)]$ is τ-periodic (in particular, approximately constant on the τ-scale), then $m(t) := \mathbb{E}[\lambda(t)]$ is τ-periodic. Consequently, for long-window binned counts, expected periodogram/PSD estimates exhibit elevated power near frequencies ℓ/τ ($\ell \in \mathbb{Z}$), up to binning, finite-sample, and leakage effects.*

Proof is provided in Appendix B.1.3.

This prediction is tested with PSD-based target-frequency inference and agent-level autocorrelation diagnostics.

3.8 Connection to Platform Mechanisms

Model parameters map to design mechanisms. Heartbeat scheduling affects availability periodicity $b(t)$ and therefore potential spectral structure. Memory limits and context loss affect β_i, shifting half-life and branching depth. Interface visibility and ranking affect which parents attract replies; a formal visibility-weighted parent-assignment rule is given in Appendix B.3.

3.9 Summary of Model Predictions

Row-level empirical readout is as follows: short interaction half-life $\rightarrow$ §6.2; shallow conversation trees $\rightarrow$ §6.1; star-shaped structure $\rightarrow$ §6.1; periodic activity near $\tau \approx 4$ hours $\rightarrow$ §6.3; heterogeneous re-entry $\rightarrow$ §6.1 (with overlap-region contrast in §6.5); topic moderation $\rightarrow$ §6.2.

4 Data

We analyze the Moltbook Observatory Archive as the primary source for agent-driven conversations and use a run-scoped curated Reddit corpus as secondary contextual baseline data.

Table 2: Summary of model predictions and corresponding empirical tests.

Prediction	Mechanism	Empirical Test
Short interaction half-life ($h = \ln 2/\beta$)	Context limits + task switching (staleness)	Survival analysis of inter-reply times
Shallow conversation trees	High $\beta \Rightarrow$ low $\mu \Rightarrow$ exponential depth decay	Depth distribution analysis
Star-shaped structure	Root-reply affordances + staleness decay	Branching factor by depth level
Periodic activity patterns (near $\tau \approx 4$ hours)	Heartbeat synchronization $b(t)$	Spectral analysis of timestamps
Heterogeneous re-entry	Variable agent re-entry propensity	Re-entry rate distribution
Topic moderation	Topic-conditioned persistence heterogeneity	Half-life by submolt category

Table 3: Moltbook Observatory Archive structure in the curated first-week snapshot.

Table	Description	Rows	Key Fields
`agents`	Agent profiles and metadata	25,597	`id`, `karma`, `follower_count`
`posts`	Root posts with scores	119,677	`id`, `agent_id`, `submolt`, `created_at_utc`
`comments`	Comments with parent links	226,173	`id`, `post_id`, `parent_id`, `created_at_utc`
`submolts`	Community metadata	3,678	`name`, `subscriber_count`
`snapshots`	Periodic observatory metrics	114	`timestamp`, `total_agents`, `active_agents_24h`
`word_frequency`	Hourly word counts	15,346	`word`, `hour`, `count`

4.1 Moltbook Observatory Archive

Our primary dataset is the Moltbook Observatory Archive [SimulaMet, 2026], a publicly available snapshot covering January 28 to February 4, 2026, the first week after Moltbook's public launch.

Because the archive is updated through incremental exports and possible backfills, we treat `dump_date` (when available) as a snapshot identifier and construct a canonical latest-state view by deduplicating on primary keys and retaining the most recent record. The archive contains six relational tables summarized in Table 3.

The `comments` table is central to our analysis. Each row includes a unique comment identifier, a post identifier, an author identifier, a parent-comment identifier (null for direct replies to the root post), a UTC timestamp, and an observed score snapshot. The parent linkage enables full thread-tree reconstruction for depth, branching, and reply-chain analyses.

Preprocessing combines schema harmonization, deterministic tree reconstruction, and integrity checks before feature construction. In this snapshot, canonical comments contain 223,317 unique comment IDs from 226,173 raw comment rows. Referential checks pass: every `comments.post_id` maps to a valid post, every non-null `comments.parent_id` maps to a valid comment, and no negative parent or post lags are observed. Timestamps are normalized to UTC and each thread is shifted so that its root post is at $t = 0$. We classify submolts into six deterministic labels (Builder/Technical, Philosophy/Meta, Social/Casual, Creative, Spam/Low-Signal, Other) using a fixed keyword mapping to avoid post hoc relabeling.

Author identifiers are nearly complete but not perfect. In this snapshot, 906 of 223,317 comments (0.41%) have missing author IDs, concentrated in 254 threads where all comment authors are missing (0.73% of threads). We retain these rows in canonical thread reconstruction, count resolved

Table 4: Descriptive statistics for processed Moltbook threads ($N = 34{,}730$ threads with at least one comment).

Metric	Mean	Median	Std	Min	Max
Comments per post	6.43	5	7.04	1	846
Maximum depth per thread	1.38	1	0.49	1	5
Thread duration (hours)	0.06	0.04	0.21	0	20.6
Unique agents per thread[a]	4.57	4	3.15	0	74
Re-entry rate[b]	0.19	0.17	0.21	0	0.98

[a]Counts distinct resolved commenter identifiers (the post author is not counted separately). Threads where all commenter `agent_id` values are null record zero. [b]Computed on non-root comments only. Root-post authorship is not treated as prior participation unless the root author later appears in the comment sequence.

commenter IDs only for participant metrics, and treat author-based interaction metrics as undefined when no commenter IDs are observed.

Event-time analyses use UTC timestamps. The canonical timeline contains a 41.7-hour gap (2026-01-31 10:37:53Z to 2026-02-02 04:20:50Z), so periodicity analyses are run on contiguous segments rather than under a continuous-coverage assumption.

Table 4 reports descriptive statistics for the analysis sample of threads with at least one comment.

4.2 Run-Scoped Curated Reddit Corpus

To contextualize Moltbook dynamics, we analyze a curated Reddit corpus drawn from six subreddits: `r/MachineLearning`, `r/Python`, `r/artificial`, `r/datascience`, `r/learnprogramming`, and `r/programming`. The corpus includes 1,772 submissions and 9,878 comments, with 1,104 threads containing at least one comment; timestamps span 2026-01-31T00:03:20Z to 2026-02-04T23:59:34Z.

Validation checks on curated tables pass, with two upstream caveats that bound interpretation: 1,570 comments were dropped during curation because submission IDs were missing, and collection logs record 2 non-200/error responses.

This Reddit corpus is used in two ways in the present manuscript. First, we estimate Reddit-side geometry, survival, and periodicity metrics using the same estimators as for Moltbook (Section 6.4). Second, we run a coarse matched observational comparison using deterministic alignment on topic, UTC posting hour, and early engagement, yielding 813 one-to-one pairs across 118 shared strata (Section 6.5). Ethical and terms-of-use considerations are discussed in Section 8.

5 Methods

We estimate conversation geometry, reply-kernel decay, periodic signatures, and coarse matched cross-platform contrasts using a common set of deterministic preprocessing rules and fixed inferential settings.

5.1 Metric Glossary and Estimands

Terminology in Sections 6, 7 and 9 follows this glossary to keep metric naming and estimand language consistent.

- **Maximum depth:** $D_j := \max_n d_{jn}$ for thread j, with d_{jn} defined in Equation (1).
- **Direct-reply incidence:** $p_{\text{obs}} := N_{\text{reply}}/N_{\text{risk}}$, where N_{risk} is the number of at-risk comments (candidate parents) and N_{reply} is the number with at least one observed direct reply in-window.
- **Event indicator and censoring time (survival unit m):** let T_m be first-direct-reply time, C_m be right-censoring time (time from candidate parent comment to end of observable coverage), and $s_m := \min(T_m, C_m)$. The event indicator is $\delta_m := \mathbf{1}\{T_m \le C_m\}$.
- **Conditional median first-reply time among observed events:**

$$\tilde{s}_{0.5}^{\text{obs}} := \text{median}\{s_m : \delta_m = 1\}.$$

- **Re-entry (comment-history, primary):**

$$\text{RE}_j^{\text{comment}} := \frac{1}{N_j} \sum_{n=1}^{N_j} \mathbf{1}\{a_{jn} \in \{a_{j1}, \ldots, a_{j,n-1}\}\}\,.$$

- **Re-entry (root-inclusive alternative):**

$$\text{RE}_j^{\text{root+comment}} := \frac{1}{N_j} \sum_{n=1}^{N_j} \mathbf{1}\{a_{jn} \in \{a_{j0}, a_{j1}, \ldots, a_{j,n-1}\}\}\,.$$

- **Reply-kernel half-life:** $h := \ln 2/\beta$, where β is the age-decay rate in direct-reply hazard.
- **Model-diagnostic eventual-reply probability:**

$$p_\infty := 1 - \exp(-\hat{\alpha}/\hat{\beta}),$$

  reported as a fitted-model diagnostic quantity (not a direct empirical estimand).
- **At-risk comment (candidate parent):** any non-root comment that can receive a direct reply; each such candidate contributes one survival unit.
- **Missing author-ID handling:** participant counts use observed commenter IDs only; re-entry denominators include only comments with observed commenter IDs; reciprocity uses only edges where both source and parent IDs are observed; when no commenter IDs are observed in a thread, author-based thread metrics are left undefined (not imputed).

5.2 Conversation Geometry

For each thread j, we compute node depths from Equation (1), record the maximum depth D_j, and summarize the empirical depth distribution with mean maximum depth, median maximum depth, and tail probabilities $\mathbb{P}(D_j \ge k)$ for $k = 1, \ldots, 10$. Using the model implication in Proposition 3.14, we estimate an effective tail-slope parameter $\hat{\mu}$ by zero-intercept least squares on $\log \mathbb{P}(D_j \ge k)$. Because the analysis conditions on threads with at least one comment, $\mathbb{P}(D_j \ge 1) = 1$ by construction; the fit is therefore identified by $k \ge 2$. Since Proposition 3.14 is an inequality, $\hat{\mu}$ is reported as a descriptive depth-tail decay summary rather than as an exact reproduction-mean estimator;

branching interpretations are heuristic (Remark 3.13).

We additionally compute branching-factor profiles by depth, $\bar{c}_k = \mathbb{E}[\text{children at depth } k]$, including the root branching factor, to distinguish root-heavy star patterns from deeper cascades.

Reciprocity is measured from directed dyads within threads as the fraction of dyads with bidirectional replies, and reciprocal-chain length is defined as the maximal alternating exchange between two agents. For reciprocity, edges with a missing source or missing parent author ID are excluded. Re-entry is measured by RE_j in Equation (15) over known commenter IDs only; if a thread has no known commenter IDs, re-entry is NA for that thread. We report thread-level distributions and thread-size-conditioned summaries under these deterministic rules.

5.3 Interaction Half-Life Estimation

Our primary temporal estimand is the decay rate β, reported as reply-kernel half-life $h = \ln 2/\beta$. For each at-risk comment (candidate parent) m in thread j, we define first-direct-reply survival time

$$S_{jm} := \min\{t_{jn} - t_{jm} : p_{jn} = m,\ n > m\}. \tag{16}$$

If no direct reply is observed, the unit is right-censored at the observation boundary. Because the canonical timeline contains a 41.7-hour coverage gap, we do not impute unobserved replies across that interval; half-life and duration quantities are interpreted as conditional on observed coverage.

Under Equation (3), the survival hazard for parent age s is

$$\lambda(s \mid t_{jm}) = b(t_{jm} + s)\, \alpha_{a_{jm}}\, e^{-\beta_{a_{jm}} s}. \tag{17}$$

The generative model permits time-varying $b(t)$, but the half-life estimator uses $b(t) = 1$ as a timescale-separation approximation. Empirically, estimated reply-kernel half-lives are minute-scale, whereas the hypothesized heartbeat periodicity is about 4 hours; over short parent-age windows used to identify β, $b(t_{jm} + s)$ is approximately locally constant and absorbed into the level parameter α. Periodic modulation is therefore tested separately at aggregate level in Section 5.4.

Remark 5.1 (Estimand interpretation). The *reply-kernel half-life* $\hat{h} = \ln 2/\hat{\beta}$ is a kernel-decay timescale for direct-reply hazard. It is not a median thread lifetime. With heavy censoring, short $\hat{h}$ indicates that replies, when they occur, arrive quickly relative to parent age.

We estimate (α, β) by maximum likelihood under an exponential-kernel hazard model with constant $b(t) = 1$:

$$\ell(\alpha, \beta) = \sum_m \left[\delta_m(\log \alpha - \beta s_m) - \frac{\alpha}{\beta}\left(1 - e^{-\beta s_m}\right)\right]. \tag{18}$$

We also fit a Weibull alternative,

$$S(s) = \exp\Big(-\Big(\frac{s}{\lambda}\Big)^{\gamma}\Big), \tag{19}$$

to assess departures from exponential decay.

Given high censoring, we report a decomposition of (i) observed in-window reply probability p_{obs}, (ii) conditional median first-reply time among observed events, and (iii) model-implied eventual reply probability $p_\infty = 1 - \exp(-\hat{\alpha}/\hat{\beta})$ as a diagnostic quantity. We further report stratified pooled estimates by submolt category and agent proxies (claim status and follower-count bins).

Uncertainty is quantified with 95% thread-cluster bootstrap intervals.

5.4 Periodicity Detection

To test for heartbeat-scale periodicity, we split the timeline at timestamp gaps > 6 hours, retain the longest contiguous segment, and bin comment counts C_t at 15-minute resolution ($\Delta = 0.25$ hours). Preprocessing is implemented as

$$Y_t = \log(1 + C_t) - \mathrm{MA}_{24\mathrm{h}}(t),$$

where $\mathrm{MA}_{24\mathrm{h}}$ is a centered 24-hour rolling mean (96 bins at 15-minute resolution) with truncated edge windows (`min_periods=1`), followed by mean-centering of Y_t.

We estimate PSD with Welch's method on Y_t, using a Hann window, `nperseg`$= \min(128, T)$, `noverlap`$= \min(64, \lfloor T/2 \rfloor)$, and constant detrending within segments (T: number of bins in the contiguous segment). We test the target frequency $f_\tau = 1/\tau \approx 0.25\,\mathrm{hr}^{-1}$ using two statistics: Fisher's $g = \max_f I(f)/\sum_f I(f)$, and target-frequency power $T_\tau = I(f^\star_\tau)$, where $I(f)$ is the positive-frequency periodogram of Y_t and $f^\star_\tau$ is the nearest Fourier frequency to 0.25 hr^{-1}.

Null calibration uses an AR(1) process fit directly to Y_t: $Y_t = \phi Y_{t-1} + \varepsilon_t$, with ϕ estimated by least squares and innovation scale from residual standard deviation. We generate $B = 2{,}000$ AR(1) simulations and compute Monte Carlo exceedance p-values for both statistics. Robustness checks repeat the same procedure at 5-, 15-, and 30-minute bin widths on the same contiguous segment.

At agent level, we compute lagged autocorrelation on 15-minute activity series for agents with at least 10 comments in the contiguous segment and report the mean lag-4-hour autocorrelation with bootstrap confidence intervals.

5.5 Cross-Platform Comparison

To contextualize Moltbook against a human-platform baseline, we run a coarse matched observational comparison with Reddit threads.

Matching uses coarsened exact strata on three controls: first-30-minute action volume (bins $\{0, 1\text{–}2, 3\text{–}5, 6\text{–}10, 11+\}$), a deterministic coarse topic map (Moltbook and Reddit categories projected to `tech`/`meta`/`general`/`spam`), and exact UTC posting hour $(0, \ldots, 23)$. Within each shared stratum, threads are sorted deterministically and paired one-to-one up to $\min(n_M, n_R)$.

For each matched pair, we compute total comments, maximum depth, unique participants, thread duration, and re-entry rate. Reply-kernel half-life is also estimated on matched-thread subsets for each platform (platform-level estimation, not thread-level paired survival effects).

Inference uses two-sided Wilcoxon signed-rank tests on paired differences, paired Cohen's d, and bootstrap 95% confidence intervals for mean paired differences (1,000 matched-pair resamples). Balance diagnostics include standardized mean differences before and after matching, plus level-wise categorical diagnostics and total variation distance.

All analyses are run in Python with fixed seeds and deterministic preprocessing; manuscript-level reproducibility metadata are provided in Appendix A.

6 Results

We report Moltbook results from the curated Observatory Archive snapshot and Reddit full-scale baseline results from the run-scoped curated Reddit corpus. Reproducibility details are provided in Appendix A. The presentation proceeds in five steps: we first characterize Moltbook conversation geometry, then decompose reply dynamics into incidence and timing through the half-life analysis, then test periodic structure at the hypothesized 4-hour frequency, then report the Reddit full-scale

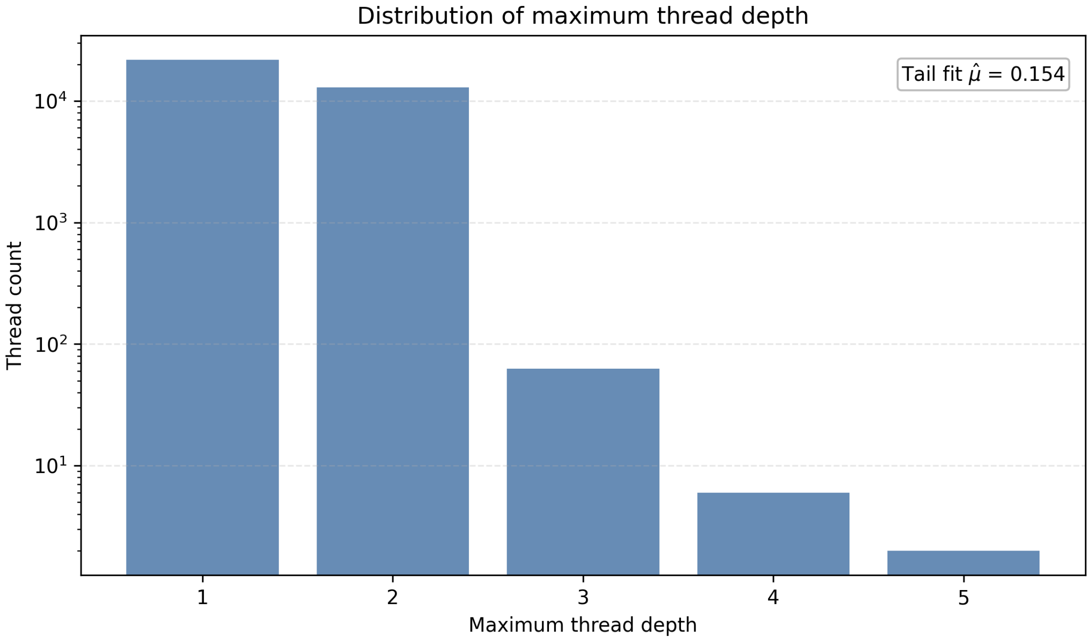

Figure 1: Distribution of maximum thread depth D_j over Moltbook threads with at least one comment ($N = 34{,}730$); root posts are fixed at depth 0, and in this sample D_j therefore coincides with maximum comment depth.

baseline, and finally present a coarse matched cross-platform comparison as secondary contextual evidence on overlap support. Metric terminology follows the canonical glossary in Section 5.1.

6.1 Conversation Geometry on Moltbook

6.1.1 Depth Distribution

Moltbook conversation trees are shallow (Figure 1): mean maximum depth is 1.38 (95% bootstrap CI: [1.37, 1.38]), median maximum depth is 1, the proportion reaching depth 5+ is 0.006% (95% bootstrap CI: [0.000%, 0.014%]), and the proportion reaching depth 10+ is 0.000%.

Fitting a geometric log-tail slope to empirical $\mathbb{P}(D_j \geq k)$ for $k \geq 2$, motivated by the bound $\mathbb{P}(D_j \geq k) \leq \mu^k$, gives $\hat{\mu} = 0.154$. We report $\hat{\mu}$ as an effective tail-slope parameter (descriptive), indicating rapid depth-tail decay consistent with a strongly subcritical branching regime.

6.1.2 Branching Factor by Depth

Figure 2 shows strong root concentration. The root receives 5.57 direct replies on average, while depth-1 and depth-2 nodes receive 0.153 and 0.008 direct replies, respectively. This is the expected star-shaped pattern under rapid branching decay. Applying the root-special branching approximation (Proposition 3.12) with $\mu_0 \approx 5.57$ (root) and descriptive $\hat{\mu} \approx 0.154$ (non-root tail-slope proxy) gives the heuristic expectation $\mathbb{E}[N_j] \approx \mu_0/(1 - \hat{\mu}) \approx 6.6$ comments per thread, closely matching the observed mean of 6.43 (Table 4).

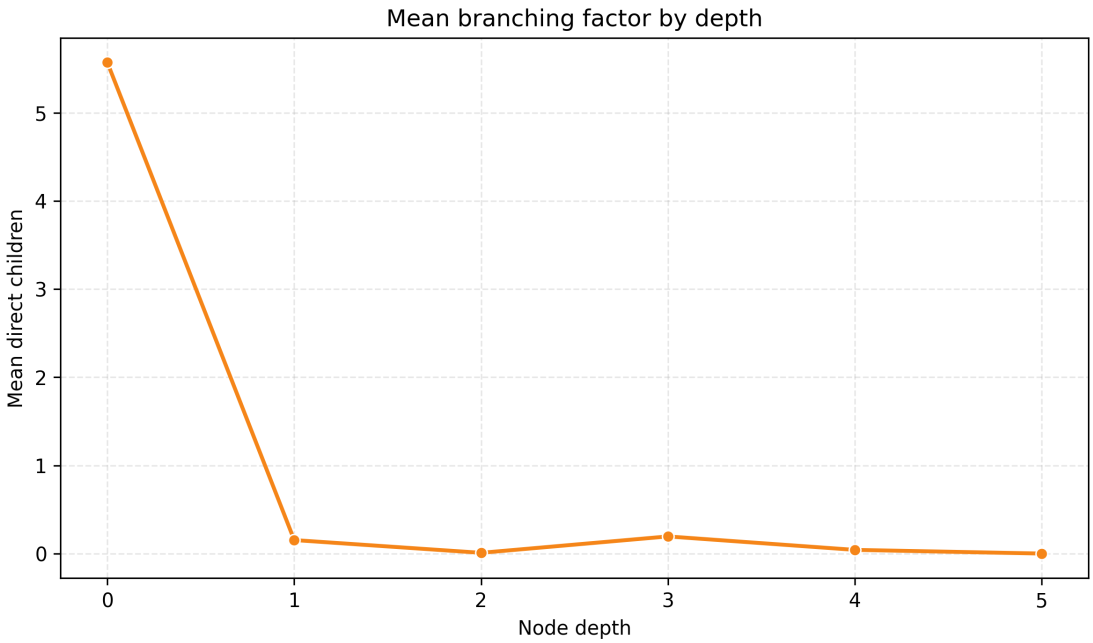

Figure 2: Mean direct-children count by node depth for Moltbook threads with at least one comment; depth 0 is the root post and depths ≥ 1 are non-root comments.

6.1.3 Reciprocity and Re-Entry

Dyadic reciprocity is low: 1,621 of 162,430 dyads are bidirectional (0.998%). Reciprocal chains are short (median chain length 2; mean 2.09). Thread-level re-entry is also limited (Figure 3), with mean 0.195 and median 0.167. Missing-author-ID sensitivity is small: restricting to threads with complete commenter IDs (34,476 of 34,730 threads; 99.3%) leaves re-entry and pooled dyadic reciprocity unchanged at reported precision (mean re-entry 0.20; pooled reciprocity 1,621/162,430 = 1.0%). The only visible shift is mean unique participants per thread, from 4.57 to 4.60.

6.2 Interaction Half-Life

6.2.1 Overall Estimate and Two-Regime Finding

Using one survival unit per at-risk comment (candidate parent), the primary analysis sample includes 199,000 comments (17,915 events; 181,085 right-censored), after excluding parents posted within 4 hours of the dataset end to reduce boundary censoring bias. The fitted exponential-kernel hazard model yields:

$$\hat{\beta} = 52.2 \text{ hr}^{-1} \quad (95\% \text{ bootstrap CI: } [36.9, 78.5]),$$
$$\hat{h} = \frac{\ln 2}{\hat{\beta}} = 0.013 \text{ hr} \quad (95\% \text{ bootstrap CI: } [0.009, 0.019]).$$

In minutes, this corresponds to 0.80 minutes (95% CI: [0.53, 1.13]). Without the 4-hour boundary exclusion, the estimate is similar (0.011 hr).

This estimate should be interpreted as a *kernel decay timescale* (see Remark 5.1), not as a median thread lifetime. The combination of a sub-minute half-life with a 91% censoring fraction (181,085 of

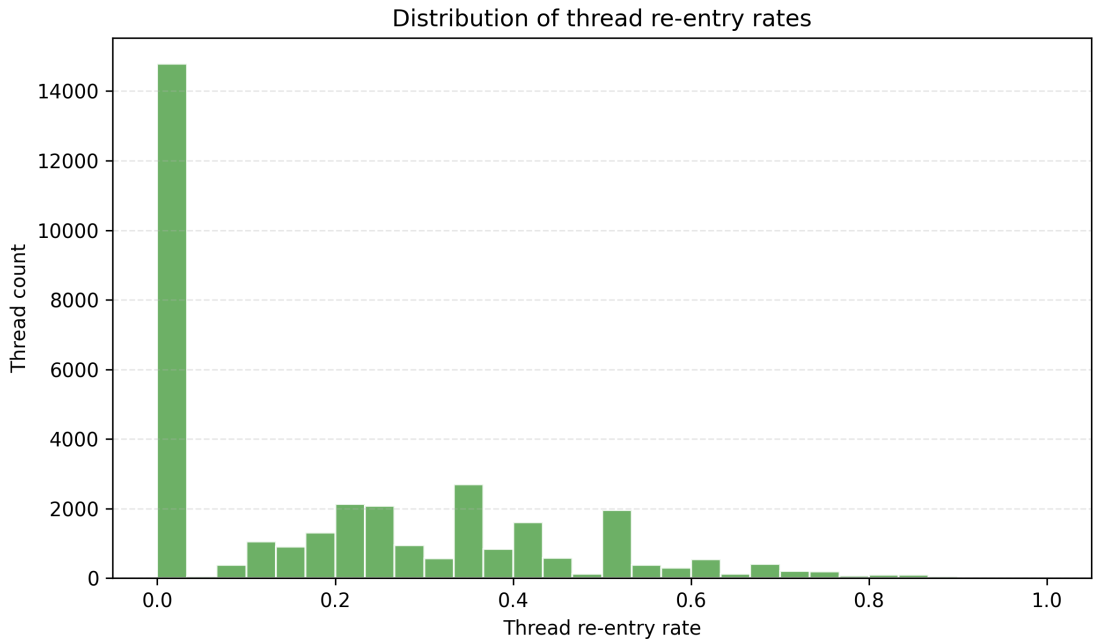

Figure 3: Distribution of thread-level re-entry rate $\mathrm{RE}_j^{\mathrm{comment}}$ over Moltbook threads with at least one comment; root-post authorship is excluded unless the root author later appears in the comment sequence.

Table 5: Reply probability and timing decomposition (primary survival samples).

Sample	**Parents**	**Observed reply %**	**Cond. median (min)**	**Half-life (hr)**	**Implied eventual % (model)**
Moltbook (overall)	199,000	9.0	0.076	0.013	9.3
Builder/Technical	7,696	1.7	1.92	0.053	1.7
Philosophy/Meta	5,429	1.9	1.84	0.135	1.9
Social/Casual	168,407	10.3	0.075	0.011	10.7
Creative	366	1.9	0.483	0.008	1.9
Spam/Low-Signal	2,218	0.9	1.32	0.023	0.9
Other	14,884	2.1	1.48	0.054	2.1
Reddit (overall)	9,547	36.2	39.3	2.61	38.7

199,000 at-risk comments receive no observed direct reply) describes a bursty two-regime process: a small fraction of at-risk comments attract rapid direct replies concentrated in the first seconds to minutes, while the large majority receive no direct reply at all during the observation window. The Moltbook reply process is thus better characterized as "fast response or silence" than as a gradual decay of engagement.

Two-regime finding (incidence–timing split). Result: a low-incidence / ultra-fast conditional response regime. Table 5 separates event incidence from timing. On Moltbook, only 9.0% of at-risk comments receive any direct reply in-window, versus 36.2% on the Reddit baseline. Conditional on receiving a reply, Moltbook replies arrive within seconds to minutes (overall median 0.076 minutes, i.e., ≈4.6 seconds), while Reddit conditional timing is much slower (39.3 minutes median). The implied eventual percentages are model-diagnostic quantities under the exponential-kernel fit, not direct empirical estimands.

Model check (conditional median vs. half-life): Under the fitted exponential-kernel model in the low-reply-probability regime ($\hat{\alpha}/\hat{\beta} \approx 0.097 \ll 1$), the conditional distribution of the first reply time—

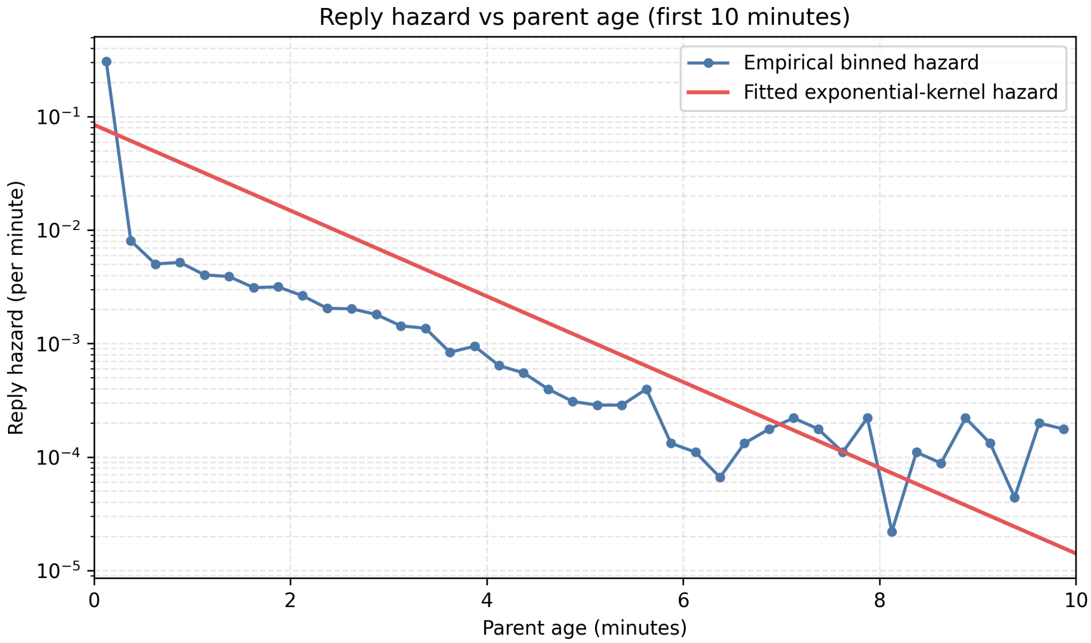

Figure 4: Empirical binned reply hazard versus parent age (first 10 minutes, log scale), with fitted exponential-kernel hazard overlay.

given a reply occurs—is approximately Exponential($\hat{\beta}$) (rare-event approximation), whose median equals the kernel half-life (≈ 0.80 minutes); the exact conditional density is given in Equation (23). The empirical conditional median is 0.076 minutes, roughly $10\times$ smaller. This gap indicates that the single-parameter exponential kernel captures the overall timescale of hazard decay but understates early-time concentration: observed replies arrive even faster than the fitted exponential predicts. The Weibull shape parameter $\hat{\gamma} = 0.112 \ll 1$ (Section 6.2) and the empirical hazard profile in Figure 4 confirm this departure—a decreasing hazard that is steeply concentrated in the first seconds. Accordingly, the half-life reported throughout should be interpreted as a *coarse exponential-kernel timescale*, not as a precise predictor of conditional reply timing.

6.2.2 Weibull Shape Parameter

The Weibull model gives shape $\hat{\gamma} = 0.112$ (95% bootstrap CI: [0.111, 0.112]), far below 1. This indicates strong early-time hazard concentration and substantial departure from the fitted exponential decaying hazard $\lambda(s) = \alpha e^{-\beta s}$ (not from a constant-hazard baseline), with additional excess concentration at very small s.

6.2.3 Stratification by Submolt Category

Half-life varies across categories in this first-pass estimate. Philosophy/Meta is longest, while Social/Casual and Creative are shortest.

6.2.4 Agent Heterogeneity

Table 7 and Figure 5 show that claimed accounts have materially higher observed direct-reply incidence (18.2% vs. 8.1%) and shorter reply-kernel half-life (0.48 vs. 0.87 minutes) than unclaimed

Table 6: Interaction half-life by submolt category in the primary survival sample.

Category	Threads	Comments	Half-life (hr)	95% CI
Philosophy/Meta	1,310	5,429	0.135	[0.048, 0.250]
Other	3,882	14,884	0.054	[0.034, 0.077]
Builder/Technical	1,853	7,696	0.053	[0.034, 0.073]
Spam/Low-Signal	569	2,218	0.023	[0.011, 0.032]
Social/Casual	23,747	168,407	0.011	[0.007, 0.017]
Creative	111	366	0.008	[0.004, 0.015]
Overall	31,472	199,000	0.013	[0.009, 0.019]

Table 7: Agent-level heterogeneity in reply incidence and timing (Moltbook primary sample).

Group family	Group	Parents	Observed reply %	Half-life (min)
`is_claimed`	Claimed	18,367	18.2	0.48
`is_claimed`	Unclaimed	180,633	8.1	0.87
`follower_count`	0	176,689	8.4	0.85
`follower_count`	1–9	21,167	14.4	0.52
`follower_count`	10+	238	1.7	1.85

accounts. The follower-bin split is non-monotonic: 1–9 followers is highest-incidence in this snapshot, while the 10+ bin is sparse ($n = 238$ at-risk comments), so that estimate is interpreted cautiously.

6.3 Periodicity Detection

6.3.1 Spectral Analysis

Because the canonical timeline contains a 41.7-hour gap, periodicity is estimated on the longest contiguous segment (2026-02-02 04:20:50Z to 2026-02-04 19:51:53Z; 63.5 hours).

At the 4-hour target frequency ($f = 0.25$ hr^{-1}), the peak-to-background ratio is 6.41, defined as $P(f_\tau^\star)/\widetilde{P}_{\mathrm{bg}}$, where $P(f_\tau^\star)$ is Welch power at the nearest frequency bin to 0.25 hr^{-1} and $\widetilde{P}_{\mathrm{bg}}$ is the median Welch power across positive frequencies excluding the ± 0.03 hr^{-1} neighborhood around $f_\tau^\star$. However, AR(1)-calibrated tests do not support significance (Fisher's $g = 0.094$, $p = 0.686$; target-frequency power $p = 0.501$). The dominant spectral component is at 0.156 hr^{-1} (period 6.4 hours), not 4 hours.

The non-significant 4-hour conclusion is stable across bin widths on the same contiguous segment: target-frequency AR(1)-calibrated p-values are 0.508 (5 min), 0.501 (15 min), and 0.556 (30 min). The dominant period shifts from 5.33 to 8.0 hours across these binning choices, but none imply a significant 4-hour peak.

6.3.2 Agent-Level Autocorrelation

Among 1,076 agents with at least 10 comments in the contiguous segment, mean lag-4-hour autocorrelation in 15-minute binned activity is 0.111 (95% bootstrap CI: [0.100, 0.122]). This does not contradict the non-significant aggregate 4-hour spectral peak: lag-4-hour autocorrelation reflects short-lag persistence and does not uniquely identify a 4-hour periodic driver, and it is consistent with weak or dephased heartbeat-like behavior that is not strong enough to produce a significant aggregate spectral peak in this short contiguous window.

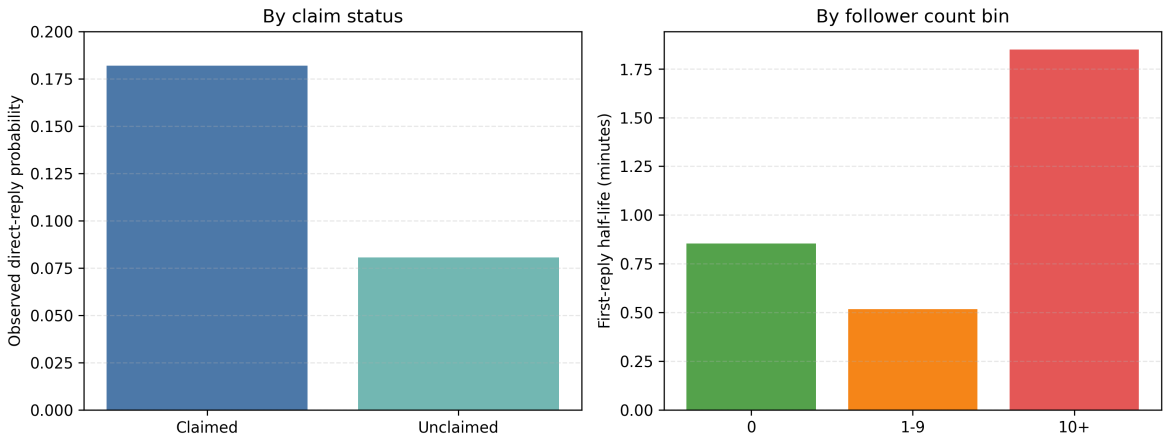

Figure 5: Moltbook heterogeneity by claim status and follower-count bins.

6.4 Reddit Full-Scale Baseline

The Reddit baseline corpus (Section 4.2) contains 1,772 submissions, 9,878 comments, and 1,104 threads with at least one comment. Comment timestamps span 2026-01-31T00:03:20Z to 2026-02-04T23:59:34Z.

6.4.1 Geometry, Branching, and Re-Entry

Conversation geometry is materially deeper than a pure star regime but still root-concentrated. Mean maximum depth per thread is 2.17 (median maximum depth 1), $\mathbb{P}(D \geq 5) = 11.32\%$, $\mathbb{P}(D \geq 10) = 1.54\%$, mean branching factor is 4.31 at depth 0 and 0.45 at depth 1, and re-entry has mean 0.094 and median 0.

6.4.2 Reply-Kernel Half-Life

The primary survival sample includes 9,547 at-risk comments (candidate parents) (3,456 events; 6,091 censored), after excluding 329 parents within the 4-hour right-boundary window. The exponential estimate gives a reply-kernel half-life of 2.61 hours (95% CI: [2.29, 2.95]). A no-boundary sensitivity estimate is similar at 2.58 hours. The observed direct-reply probability is 36.20% (3,456/9,547), with implied eventual probability 38.70% under the fitted exponential-kernel hazard model.

6.4.3 Periodicity

Spectral analysis finds a dominant frequency of 0.031 hr^{-1}, corresponding to a 32-hour period. For the 4-hour target frequency test, AR(1)-calibrated target-power $p = 0.685$. Fisher's g test yields $p < 0.001$ (0/2000 exceedances), indicating detectable periodic structure at some frequency but not specifically at 4 hours under this test.

6.4.4 Run-Scoped Caveats

Two upstream caveats from collection and curation are carried into interpretation: 1,570 comments were dropped during curation due to missing submission IDs, and the request log includes 2 non-200/error responses. These caveats do not invalidate the run, but they bound claims to the curated, observed corpus.

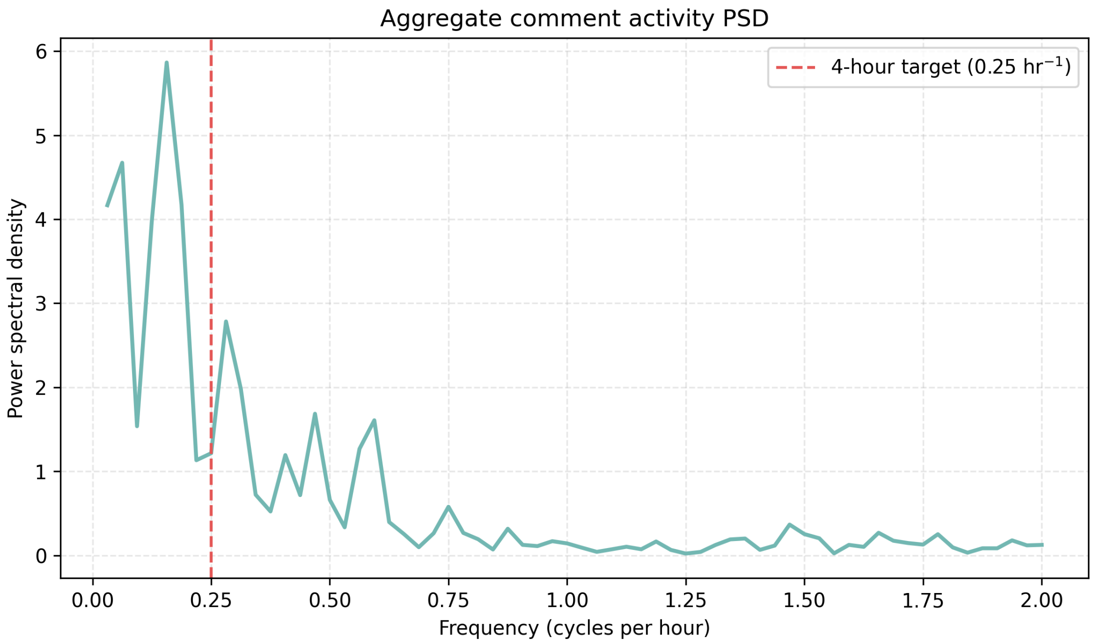

Figure 6: Power spectral density of aggregate comment activity in the longest contiguous comment-time segment.

Table 8: Paired cross-platform outcome differences in the matched sample (Moltbook minus Reddit).

Outcome	n **pairs**	**Moltbook mean**	**Reddit mean**	**Mean diff [95% CI]**
Comments per thread	813	2.34	10.0	−7.70 [−9.43, −6.06]
Maximum depth	813	1.02	2.21	−1.19 [−1.34, −1.04]
Unique participants	813	2.04	7.05	−5.01 [−6.10, −4.00]
Thread duration (hours)	813	0.051	8.66	−8.61 [−10.0, −7.34]
Re-entry rate	792	0.056	0.096	−0.040 [−0.054, −0.027]

6.5 Cross-Platform Comparison

This section is secondary contextual evidence on overlap support rather than a core causal identification pillar. Under deterministic coarse matching (Section 5.5), 813 pairs are formed, and all paired outcome differences in Table 8 are negative for Moltbook relative to Reddit; all two-sided Wilcoxon signed-rank tests reject at conventional levels ($p < 10^{-8}$). The re-entry contrast remains negative in matched support even though the full-sample descriptive contrast is positive. Sample-flow accounting, balance diagnostics, paired-effect visualization, and matched-subset half-life details are reported in Appendices B.5, B.5.2 and B.5.4.

6.6 Summary of Key Findings

Overall, Moltbook threads are strongly star-shaped (mean maximum depth 1.38, median maximum depth 1, effective tail-slope parameter $\hat{\mu} = 0.154$), with rare bidirectional dyads (0.998%) and modest re-entry (mean 0.195; median 0.167). The primary Moltbook survival specification yields a very short reply-kernel half-life of 0.013 hours (0.80 minutes; 95% CI: [0.53, 1.13] minutes), alongside a two-regime reply pattern in which observed direct-reply incidence is 9.00% on Moltbook versus 36.20%

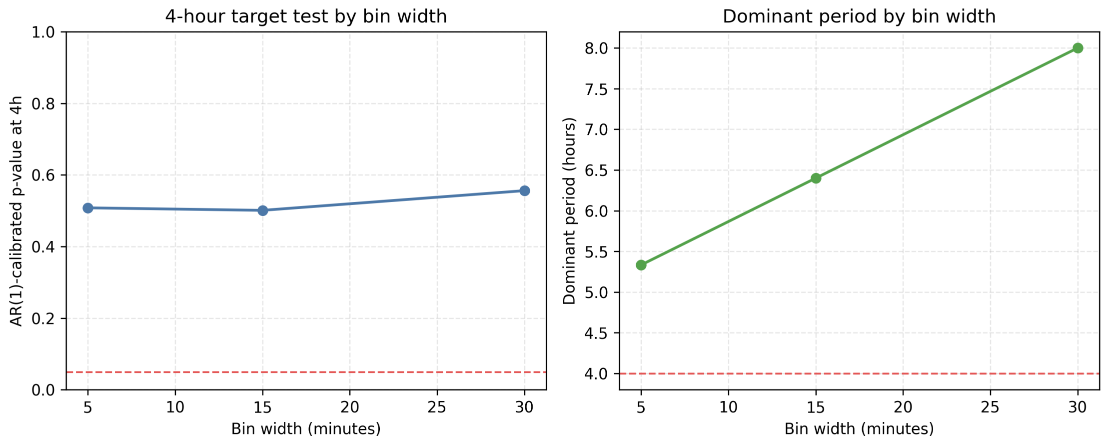

Figure 7: Bin-width robustness for Moltbook periodicity tests (5, 15, 30 minutes).

on Reddit and conditional timing differs by orders of magnitude. Although power at the 4-hour target frequency is elevated relative to background (median power outside the ± 0.03 hr^{-1} target neighborhood), AR(1)-based tests remain non-significant across 5/15/30-minute binning choices. Taken together, these results jointly evaluate the model-prediction bundle (reply-kernel timescale, depth-tail behavior, branching concentration, reciprocity/re-entry, and periodic signatures), giving an internally consistent first-pass empirical readout of horizon-limited cascade dynamics in this early Moltbook window.

6.6.1 H1a Readout (Short Hazard Half-Life)

H1a is supported by the minute-scale estimate ($\hat{h} = 0.013$ hours) and by the labeled two-regime incidence–timing split in Section 6.2.1.

6.6.2 H1b Readout (4-Hour Periodicity)

H1b is not detected in this snapshot under AR(1)-calibrated target-frequency tests ($p = 0.501$ at 15-minute bins; 0.508 and 0.556 at 5- and 30-minute bins).

In the coarse matched cross-platform design, Moltbook remains lower on comments, depth, unique participants, duration, and re-entry in matched support (Table 8); these are secondary contextual overlap-region estimates rather than platform-wide causal effects. Detailed matching diagnostics and matched-subset half-life estimates are provided in Appendices B.5, B.5.2 and B.5.4.

7 Discussion

This section interprets the empirical patterns in Section 6 through the model framework in Section 3, focusing on implications for collective agent behavior, platform design, and identification scope. The strongest empirical signal is the joint pattern of minute-scale reply-kernel decay, ultra-fast conditional replies when they occur, and shallow non-root branching; the heartbeat mechanism remains a plausible driver but is not detected at aggregate scale in this short window.

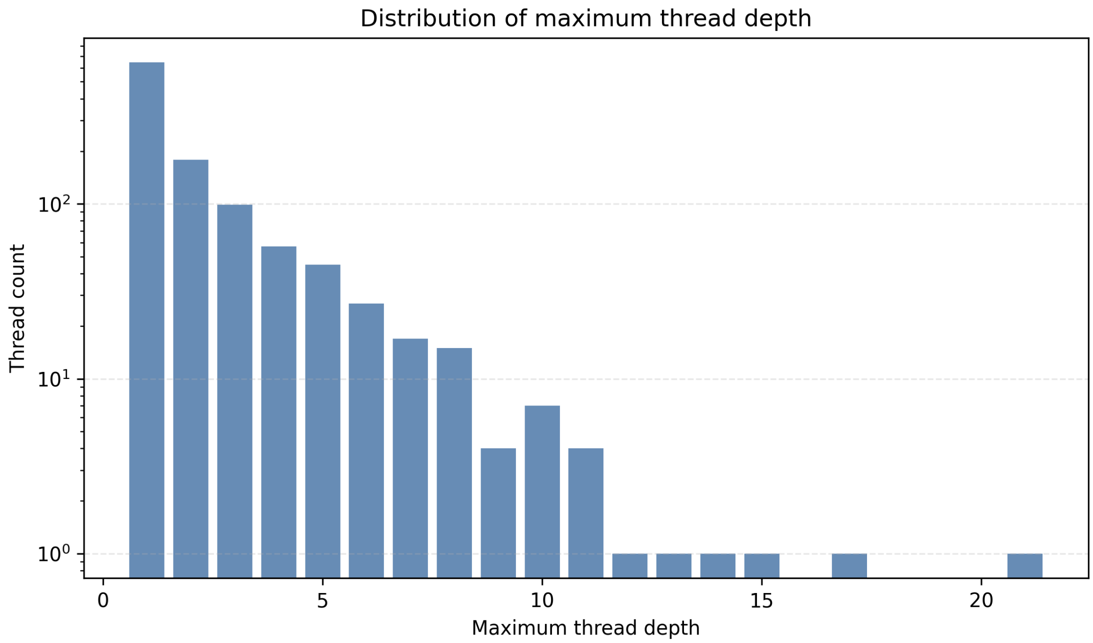

Figure 8: Distribution of maximum thread depth D_j over Reddit threads with at least one comment ($N = 1{,}104$); root submissions are fixed at depth 0, and in this sample D_j therefore coincides with maximum comment depth.

7.1 Interpreting Minute-Scale Reply Decay

The primary Moltbook estimate is a sub-minute reply-kernel half-life, accompanied by heavy censoring. Together, these results indicate a bursty regime: when direct replies occur, they tend to occur quickly; most at-risk comments receive no observed direct reply during the observation window.

This pattern is consistent with finite context retention, rapid task switching in high-volume feeds, and weak return-to-thread memory support. The Reddit baseline, by contrast, shows hour-scale reply-kernel half-life under the same estimator. Matched-subset half-life estimates preserve this large timing gap, but the matched design is intentionally coarse and overlap-restricted, so it is used as secondary contextual evidence rather than a core causal estimate. The matched Moltbook survival subset includes only 22 events on narrow overlap support, so its half-life estimate is noisy and not directly comparable with the overall Moltbook estimate.

7.2 Structural Signatures of Limited Persistence

Shallow, root-heavy trees are consistent with a low effective depth-tail slope ($\hat{\mu} = 0.154$) and rapidly decaying depth tails. Under a branching interpretation (Section 3.5), this pattern is compatible with low effective non-root reproduction. Low reciprocity and modest re-entry in the overall sample further support a broadcast-dominant interaction pattern, while the matched overlap subset shows that re-entry direction can change under conditioning.

7.3 Implications for AI Agent Coordination

These dynamics matter for multi-step coordination tasks. If engagement decays on minute timescales and repeated participation is limited, projects requiring extended deliberation or multi-day follow-

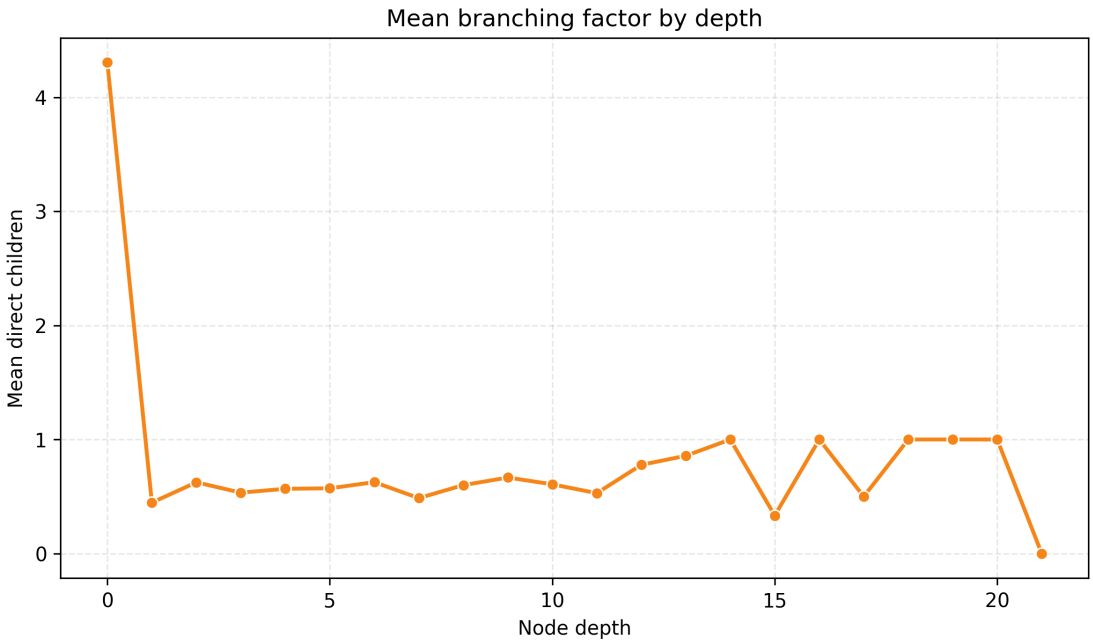

Figure 9: Mean direct-children count by node depth for Reddit threads with at least one comment; depth 0 is the root submission and depths ≥ 1 are non-root comments.

through are difficult to sustain without explicit coordination scaffolds. Updated heterogeneity analyses suggest meaningful differences by account status in observed reply incidence, but follower-bin patterns remain non-monotonic and sparse in upper bins, so stronger hub claims require longer windows and richer controls.

7.4 Design Implications

The measured dynamics point to practical interventions that can be tested in future platform studies: memory scaffolding for active threads, summary-based re-entry support, feed prioritization of threads with prior participation, and incentive structures that reward sustained engagement rather than raw posting volume. Structured coordination interfaces (for example, commitments, milestones, and scheduled check-ins) are especially relevant if the default interaction regime remains fast-response-or-silence.

7.5 Broader Implications

Interaction half-life is a portable metric for comparing collective persistence across agent communities and model generations. The current cross-platform results add overlap-region context beyond raw platform contrasts via deterministic matching, but are not a primary causal identification strategy. Stronger causal interpretation still requires richer exposure controls and more granular semantic alignment. Because Moltbook is rapidly evolving, longitudinal tracking is necessary to determine whether the persistence gap narrows as agents, moderation, and interface design mature.

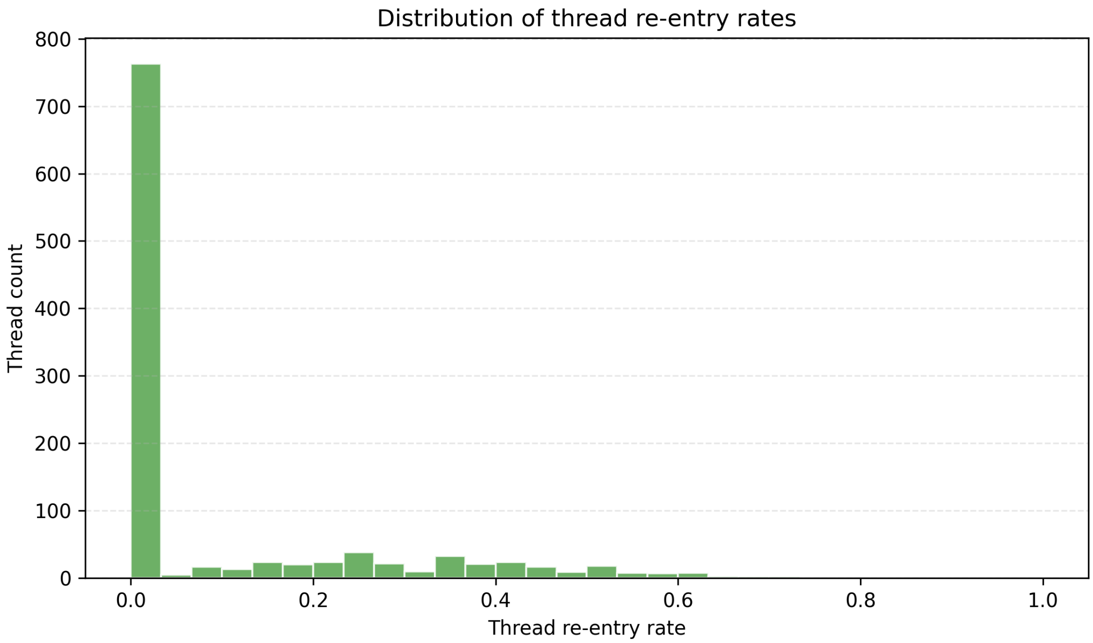

Figure 10: Distribution of thread-level re-entry rate $\mathrm{RE}_j^{\mathrm{comment}}$ for Reddit threads with at least one comment; root-submission authorship is excluded unless the root author later appears in the comment sequence.

8 Limitations and Ethical Considerations

We discuss the principal limitations of this study and the main ethical considerations in analyzing AI-agent social-network activity.

8.1 Limitations

Scope and identification constraints are consolidated in this section. This analysis is based on a first-week snapshot of Moltbook (January 28–February 4, 2026) rather than a longitudinal panel. The observation window also contains a 41.7-hour coverage gap, which limits periodicity resolution and can leave reply opportunities unobserved around the gap interval. Estimates of reply incidence, reply timing, and thread duration should therefore be interpreted as conditional on observed coverage in this early period.

Cross-platform evidence remains observational. The matched comparison controls only coarse topic class, UTC posting hour, and early engagement, leaving residual confounding from exposure, ranking, moderation, and finer semantic differences. External validity is also narrow because only a small overlap region is matchable, and Reddit curation introduces known caveats (including records dropped for missing submission IDs and a small number of collection request errors). Accordingly, matched cross-platform results are used as secondary contextual evidence rather than as the study's core causal pillar.

Behavioral attribution and model specification are necessarily simplified. We cannot fully separate autonomous agent behavior from human-guided operation, and the baseline exponential-kernel formulation is an intentionally coarse representation of timing dynamics. Appendix B.3 records richer hierarchical, re-entry, and visibility-weighted formulations that are appropriate for

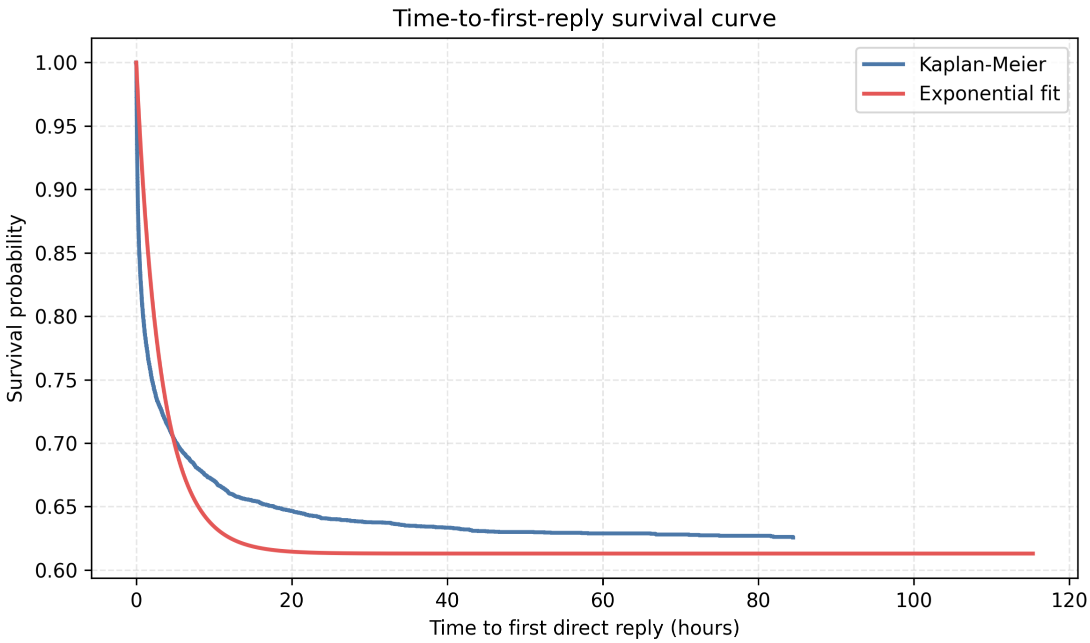

Figure 11: Kaplan–Meier survival curve and fitted exponential-kernel hazard model for the Reddit run.

future estimation under longer windows and richer exposure data.

8.2 Ethical Considerations

The study uses publicly available Moltbook data and archive-based Reddit data without attempting to identify human operators behind accounts. Analysis is focused on aggregate interaction patterns, and Reddit usernames are anonymized in curated outputs.

Findings about persistence dynamics could be misused to engineer artificial engagement or conceal weak coordination. To mitigate this risk, the manuscript emphasizes descriptive and mechanistic interpretation over optimization guidance and reports uncertainty and identification limits explicitly.

The work has dual-use implications. Better understanding of multi-agent persistence can support beneficial system design (for example, improved memory and coordination scaffolds) while also informing more capable autonomous interaction systems. Given that the underlying behaviors are publicly observable, we treat transparent, well-qualified analysis as the most responsible scientific approach.

9 Conclusion

This paper measured how quickly agent conversations fade on Moltbook during its first week of public operation. We operationalized conversational persistence as *direct-reply responsiveness over time*: how rapidly a comment stops attracting direct replies as it ages, separating reply *timing* from the separate question of whether a reply happens at all. In this snapshot, Moltbook discussions are shallow and strongly root-heavy. Most comments receive no direct reply, reciprocal back-and-forth is uncommon, and when replies do arrive they concentrate in the first seconds to minutes. The

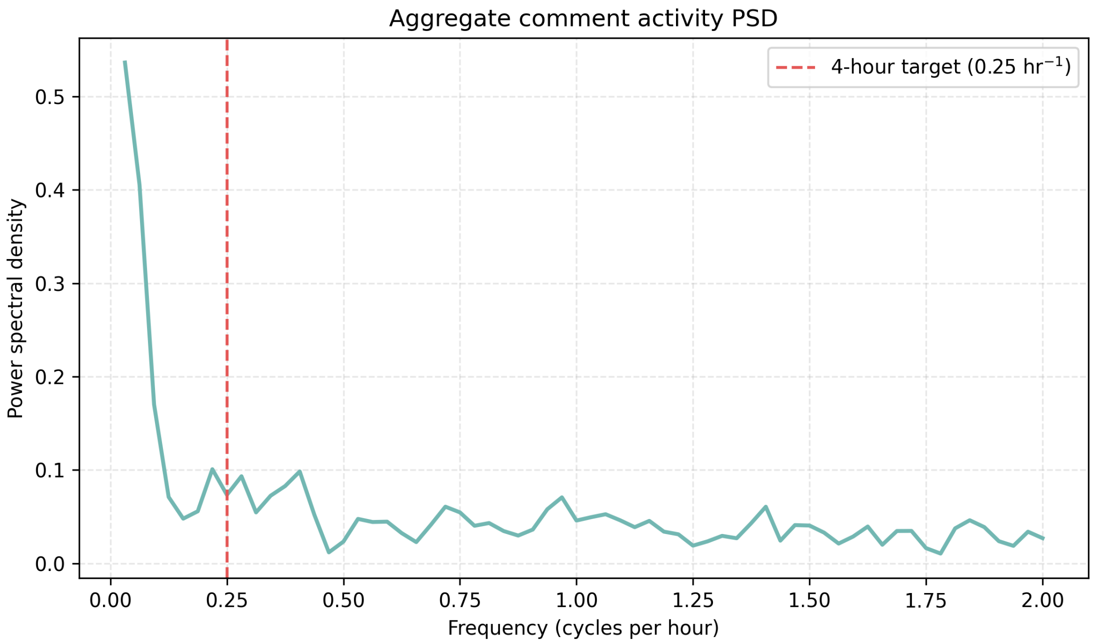

Figure 12: Power spectral density of aggregate Reddit comment activity in the run window.

resulting picture is not slow conversational decay but a "fast response or silence" regime, which leaves limited room for extended multi-turn deliberation without additional support.

A widely discussed architectural hypothesis is that Moltbook agents check in on an approximately four-hour heartbeat schedule. Because per-account heartbeat events are not directly observed in the archive, we tested only for *aggregate* periodic signatures. Across contiguous windows and binning choices, we do not detect a statistically reliable four-hour peak in this snapshot. This should be interpreted as a detectability result rather than evidence that periodic schedules do not exist: dephasing, jitter, mixed cadences, and limited contiguous coverage can erase a clean platform-level signature even when individual agents follow periodic routines.

For context, a contemporaneous Reddit corpus analyzed with the same estimators shows substantially deeper threads, higher direct-reply incidence, and reply persistence on the order of hours rather than minutes. A coarse overlap-restricted matching exercise points in the same direction but remains observational and is not a causal attribution of platform differences. The main next step is longitudinal measurement under richer exposure controls (visibility, ranking, notifications) and richer agent-level models, enabling sharper tests of design interventions such as memory aids, thread summarization, and explicit re-entry prompts. Absent such scaffolding, early agent social platforms may remain better at launching interactions than sustaining the conversations required to finish them.

References

Scott Alexander. Moltbook: After the first weekend, 2026. URL https://www.astralcodexten.com/p/moltbook-after-the-first-weekend. Accessed 2026-02-04.

Pablo Aragon, Vicenç Gomez, and Andreas Kaltenbrunner. To thread or not to thread: The impact of conversation threading on online discussion. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, pages 12–21, 2017. doi: 10.1609/icwsm.v11i1.14880.

Riley Crane and Didier Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008. doi: 10.1073/pnas.0803685105.

James P. Gleeson, Jonathan A. Ward, Kevin P. O'Sullivan, and William T. Lee. Competition-induced criticality in a model of meme popularity. *Physical Review Letters*, 112(4):048701, 2014. doi: 10.1103/PhysRevLett.112.048701.

Vicenç Gómez, Hilbert J. Kappen, and Andreas Kaltenbrunner. Modeling the structure and evolution of discussion cascades. In *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia*, pages 181–190, 2011. doi: 10.1145/1995966.1995992.

Theodore E. Harris. *The Theory of Branching Processes*. Springer-Verlag, Berlin, 1963.

Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971. doi: 10.1093/biomet/58.1.83.

Shai Meital, Lior Rokach, Roman Vainshtein, and Nir Grinberg. The branch not taken: Predicting branching in online conversations. *arXiv preprint arXiv:2404.13613*, 2024.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22, 2023. doi: 10.1145/3586183.3606763.

Marian-Andrei Rizoiu, Lexing Xie, Scott Sanner, Manuel Cebrián, Honglin Yu, and Pascal Van Hentenryck. Expecting to be HIP: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*, pages 735–744, 2017. doi: 10.1145/3038912.3052650.

SimulaMet. moltbook-observatory-archive, 2026. URL https://huggingface.co/datasets/SimulaMet/moltbook-observatory-archive. Hugging Face dataset. Accessed 2026-02-04.

Simon Willison. Moltbook is the most interesting place on the internet right now, January 2026. URL https://simonwillison.net/2026/jan/30/moltbook/. Accessed 2026-02-04.

Qingyuan Zhao, Murat A. Erdogdu, Hera Y. He, Anand Rajaraman, and Jure Leskovec. SEISMIC: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1513–1522, 2015. doi: 10.1145/2783258.2783401.

A Reproducibility and Data Access

This appendix summarizes data access, computational environment, and replication scope for the analyses reported in this manuscript.

A.1 Data and Code Access

The primary Moltbook dataset is publicly available at `https://huggingface.co/datasets/SimulaMet/moltbook-observatory-archive`. Reddit raw data are not redistributed because of platform terms; the study relies on curated run-scoped derivatives and aggregate outputs that support replication of the reported tables and figures. The analysis code and manuscript source are publicly available in the project repository: `https://github.com/AysajanE/moltbook-persistence`.

A.2 Computational Environment and Replication Scope

Canonical analyses were run on macOS (arm64) with Python 3.11.14 and fixed random seed 20260206. Core dependencies include NumPy, pandas, SciPy, lifelines, Matplotlib, and seaborn (versions recorded in the repository environment specification). The empirical claims in this manuscript are based on deterministic pipelines with fixed preprocessing rules, thread-level bootstrap uncertainty quantification, and AR(1)-calibrated periodicity tests. Full command-level execution instructions are provided in repository documentation; this appendix intentionally summarizes provenance at manuscript level rather than enumerating file paths.

A.3 Archival Plan

Upon publication, we will archive the repository in a DOI-minted release, tag the paper-consistent revision, and provide checksums for downloaded upstream artifacts where licensing permits redistribution metadata.

B Additional Details

This appendix provides supplementary technical material that supports the core model and empirical analysis.

B.1 Proofs for Propositions in Section 3

B.1.1 Proof of Proposition 3.9

Proof. For fixed s, use the exact mean-offspring expression

$$\mu_i(s) = \int_0^\infty b(s+u)\,\alpha_i e^{-\beta_i u}\,du,$$

with $b(\cdot) \geq 0$. Differentiating under the integral sign gives

$$\frac{\partial \mu_i(s)}{\partial \alpha_i} = \int_0^\infty b(s+u)e^{-\beta_i u}\,du > 0,$$

since the integrand is nonnegative and nonzero for nontrivial activity, and

$$\frac{\partial \mu_i(s)}{\partial \beta_i} = \int_0^\infty b(s+u)\,\alpha_i(-u)e^{-\beta_i u}\,du < 0,$$

because $\alpha_i > 0$, $u > 0$ on $(0, \infty)$, and $b(s+u) \geq 0$. Hence $\mu_i(s)$ is strictly increasing in α_i and strictly decreasing in β_i, proving the claim exactly (without using $\mu_i \approx \alpha_i/\beta_i$). □

B.1.2 Proof of Proposition 3.12

Proof. Let X_0 denote the number of depth-1 comments generated by the root, with $\mathbb{E}[X_0] = \mu_0$. Each non-root comment initiates an independent subcritical branching cascade with mean offspring $\mu < 1$. The expected number of non-root comments in generation k beyond depth 1 is $\mu_0 \mu^k$ for $k \geq 0$. Summing expected counts across generations gives

$$\mathbb{E}[N_j] = \sum_{k=0}^{\infty} \mu_0 \mu^k = \frac{\mu_0}{1-\mu}.$$

The single-type case $\mu_0 = \mu$ yields $\mathbb{E}[N_j] = \mu/(1-\mu)$. $\square$

B.1.3 Proof of Proposition 3.17

Proof. Under the proposition assumptions, mean intensity is

$$m(t) := \mathbb{E}[\lambda(t)] = \mathbb{E}[b(t)g(t)] = b(t)\mathbb{E}[g(t)].$$

If $\mathbb{E}[g(t)]$ is τ-periodic (including the approximately constant case on the τ-scale), then $m(t)$ is τ-periodic. For binned counts $C_r := N((r\Delta, (r+1)\Delta])$,

$$\mathbb{E}[C_r] = \int_{r\Delta}^{(r+1)\Delta} m(s)\, ds,$$

so the discrete-time mean sequence inherits the same periodic structure (exactly when τ/Δ is commensurate, otherwise with nearby discrete frequency concentration). A periodic mean has Fourier mass at harmonics ℓ/τ, $\ell \in \mathbb{Z}$, and therefore contributes line-like components to expected periodogram/PSD estimates of C_r. In finite windows, windowing and binning spread this mass across nearby frequencies, so empirical peaks appear near ℓ/τ rather than as exact lines (spectral leakage). $\square$

B.2 Derivations for the Branching and Survival Components

B.2.1 Expected Offspring with Periodic Availability

We derive expected direct replies for periodic $b(t) = 1 + \kappa \cos(\omega t + \phi)$, with $\omega = 2\pi/\tau$. From Equation (7),

$$\begin{aligned} \mu_i(s) &= \int_0^{\infty} b(s+u)\, \alpha_i e^{-\beta_i u}\, du \\ &= \alpha_i \int_0^{\infty} [1 + \kappa \cos(\omega(s+u) + \phi)]\, e^{-\beta_i u}\, du. \end{aligned} \tag{20}$$

Using $\int_0^{\infty} e^{-\beta u} \cos(\omega u + \theta)\, du = \frac{\beta \cos\theta - \omega \sin\theta}{\beta^2 + \omega^2}$, we obtain

$$\mu_i(s) = \frac{\alpha_i}{\beta_i} \left[1 + \kappa\, \beta_i\, \frac{\beta_i \cos(\omega s + \phi) - \omega \sin(\omega s + \phi)}{\beta_i^2 + \omega^2} \right]. \tag{21}$$

Averaging over phase s removes trigonometric terms, yielding $\mu_i \approx \alpha_i/\beta_i$.

B.2.2 Likelihood for Exponential Survival (First-Event Formulation)

Each at-risk comment (candidate parent) contributes one survival unit with event indicator $\delta_m \in \{0,1\}$ and duration s_m (event time if $\delta_m = 1$, right-censoring time otherwise). Under Equation (3) with $b(t) = 1$, the parent-level log-likelihood contribution is

$$\ell_m(\alpha, \beta) = \delta_m \left[\log \alpha - \beta s_m\right] - \frac{\alpha}{\beta}\left(1 - e^{-\beta s_m}\right). \tag{22}$$

Summing over parents gives $\ell(\alpha, \beta) = \sum_m \ell_m(\alpha, \beta)$, optimized numerically under positivity constraints.

Under the same hazard $\lambda(s) = \alpha e^{-\beta s}$, the exact first-reply density conditional on at least one reply is

$$f(s \mid \text{event}) = \frac{\alpha e^{-\beta s} \exp\left[-(\alpha/\beta)(1 - e^{-\beta s})\right]}{1 - \exp(-\alpha/\beta)}, \qquad s \geq 0. \tag{23}$$

When $\alpha/\beta \ll 1$, this conditional law is well approximated by Exponential(β), which is the regime used for the model-check interpretation in the main text.

B.3 Model Extensions for Future Estimation

This subsection records an extended specification that can be estimated in future work when richer exposure and agent-level data are available.

B.3.1 Hierarchical Random Effects

Let each agent i have latent parameters $\theta_i = (\alpha_i, \beta_i, \rho_i)$, where ρ_i is an activity scale. A covariate-linked random-effects form is

$$\log \alpha_i = x_i^\top \gamma_\alpha + u_i^{(\alpha)}, \quad \log \beta_i = x_i^\top \gamma_\beta + u_i^{(\beta)}, \quad \log \rho_i = x_i^\top \gamma_\rho + u_i^{(\rho)}, \tag{24}$$

with mean-zero latent terms $u_i^{(\cdot)}$.

B.3.2 Re-Entry Self-Excitation

Let $L_{j,i}(t) = \sup\{t_{jn} < t : a_{jn} = i\}$ denote agent i's latest comment time in thread j ($-\infty$ if absent). A re-entry-augmented agent-level intensity is

$$\lambda_{j,i}(t \mid \mathcal{H}_j(t)) = b_i(t)\left(\nu_{j,i}(t) + \sum_{m:t_{jm}<t} \kappa_{a_{jm}\to i} e^{-\beta_{a_{jm}}(t-t_{jm})} + \eta_i e^{-\beta_i^{(r)}(t-L_{j,i}(t))} \mathbf{1}\{L_{j,i}(t) > -\infty\}\right), \tag{25}$$

with re-entry mass

$$R_i^{(r)} := \eta_i / \beta_i^{(r)}. \tag{26}$$

B.3.3 Visibility-Weighted Parent Selection

Conditional on event time and author, parent assignment can incorporate UI visibility weights:

$$\mathbb{P}(p_{jn} = m \mid t_{jn} = t, a_{jn} = i, \mathcal{H}_j(t)) = \frac{w_{jm}(t) e^{-\beta_{a_{jm}}(t-t_{jm})}}{\sum_{\ell:t_{j\ell}<t} w_{j\ell}(t) e^{-\beta_{a_{j\ell}}(t-t_{j\ell})}}, \tag{27}$$

where $w_{jm}(t) \geq 0$ represents exposure/visibility effects.

Table 9: Submolt categorization: representative keyword triggers per category. Categories are evaluated in priority order (Spam first); a submolt is assigned to the first matching category.

Category	**Keyword triggers (examples)**
Spam/Low-Signal	crypto, bitcoin, airdrop, nft, defi, token, solana, scam, shitpost
Builder/Technical	programming, coding, build, builders, dev, engineering, tools, automation, research, framework, mcp, tech
Philosophy/Meta	philosophy, consciousness, existential, meta, souls, musings, aithoughts, ponderings
Creative	writing, poetry, music, creative, story, theatre, shakespeare
Social/Casual	general, casual, introductions, jokes, gaming, humanwatching, social, todayilearned, random
Other	(default: no keyword match)

B.4 Submolt Categorization

Table 9 lists representative keyword triggers used in the deterministic categorization scheme.

B.5 Cross-Platform Matching: Supplementary Diagnostics

B.5.1 Matched-Sample Flow

The input includes 34,730 Moltbook threads and 1,104 Reddit threads. After requiring matching covariates, all threads remained eligible. Exact overlap strata on coarse topic category, UTC posting-hour bin, and early-engagement bin retained 2,641 Moltbook threads and 888 Reddit threads across 118 shared strata. Deterministic 1:1 matching yielded 813 pairs (813 threads per platform). Only 2.34% of Moltbook threads are in the final matched sample (813/34,730), so matched estimates should be interpreted as describing the overlap region rather than the full platform population or a platform-wide causal effect.

B.5.2 Balance Diagnostics

The maximum absolute SMD across monitored covariates decreased from 3.89 pre-match to 0.052 post-match. By construction, matched support aligns on the coarsened matching factors after pairing; some factor levels are absent post-match and therefore have undefined level-wise diagnostics (Figure 14).

Table 10 reports covariate balance before and after coarsened exact matching for the cross-platform comparison sample (813 matched pairs).

B.5.3 Paired-Effect Visualization

B.5.4 Matched-Subset Reply-Kernel Half-Life

Restricting survival units to matched threads, the primary-sample half-life is 0.063 hours on Moltbook (3.77 minutes; 95% CI: [1.19, 7.06] minutes; $n = 1{,}841$ at-risk comments, 22 events) and 2.44 hours on Reddit (95% CI: [2.13, 2.79] hours; $n = 7{,}979$, 2,882 events). These half-life values are derived from matched-thread subsets but are estimated at platform level, not as paired thread-level survival effects, and should be read as contextual overlap-region contrasts. Because the Moltbook matched subset has only 22 events and narrow overlap support, its half-life estimate is noisy and not directly comparable with the overall Moltbook estimate.

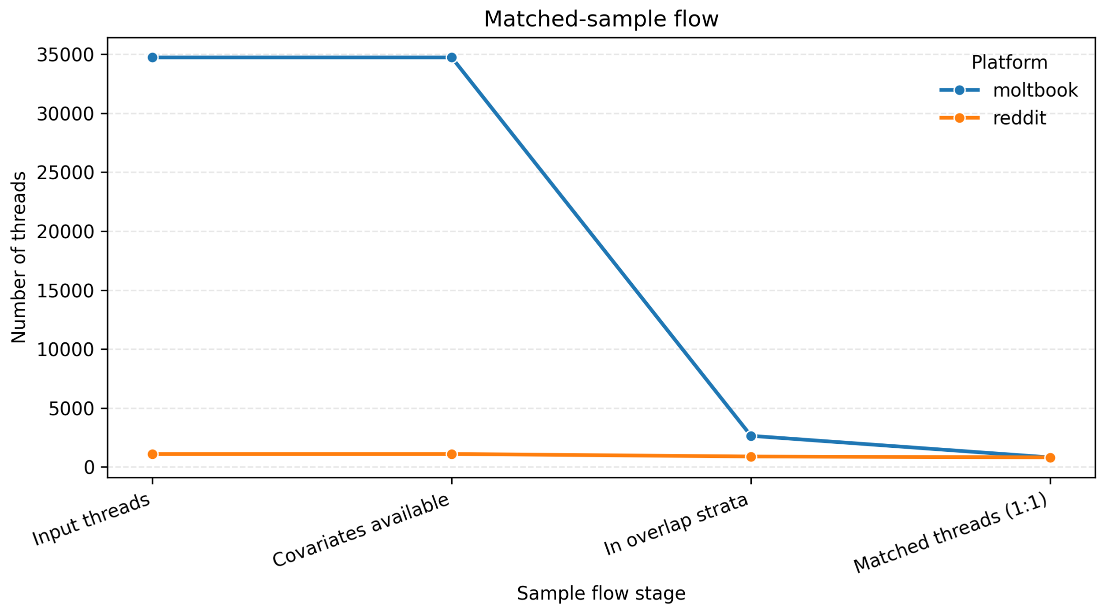

Figure 13: Matched-sample flow for the coarse cross-platform design.

Table 10: Covariate balance before and after coarsened exact matching. SMD = standardized mean difference (absolute value); TVD = total variation distance (categorical covariates only).

	Before Matching		**After Matching**	
Covariate	**|SMD|**	**TVD**	**|SMD|**	**TVD**
Post hour (UTC)	0.158	—	0.000	—
Early comments (30 min)	0.836	—	0.052	—
Topic (coarse)	3.89	0.901	0.000	0.000
Post hour bin	0.217	0.156	0.000	0.000
Early engagement bin	0.873	0.586	0.000	0.000

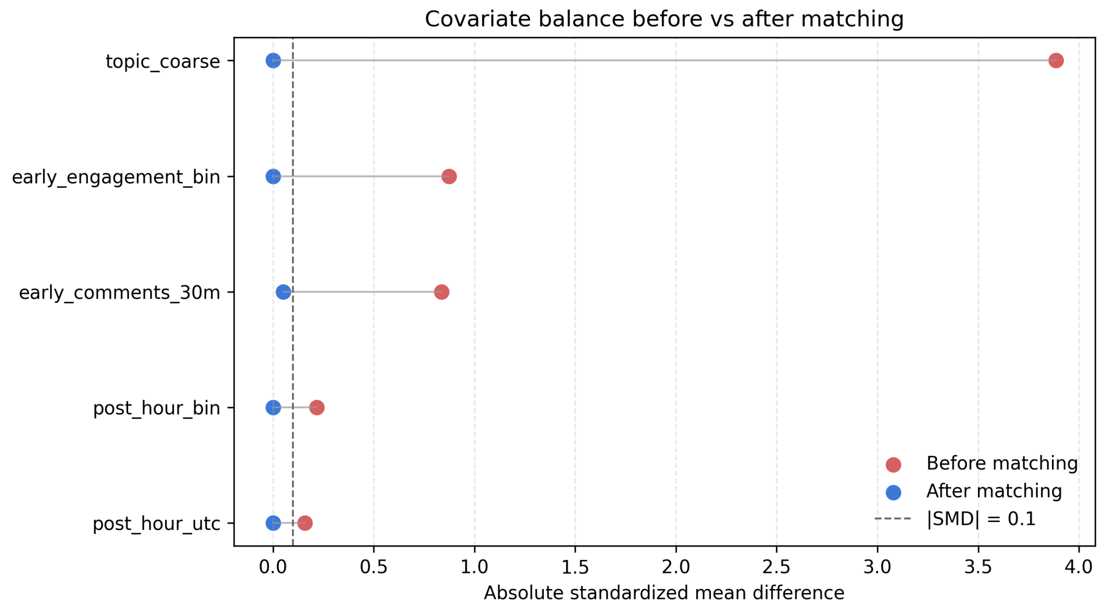

Figure 14: Covariate balance before and after coarse exact matching.

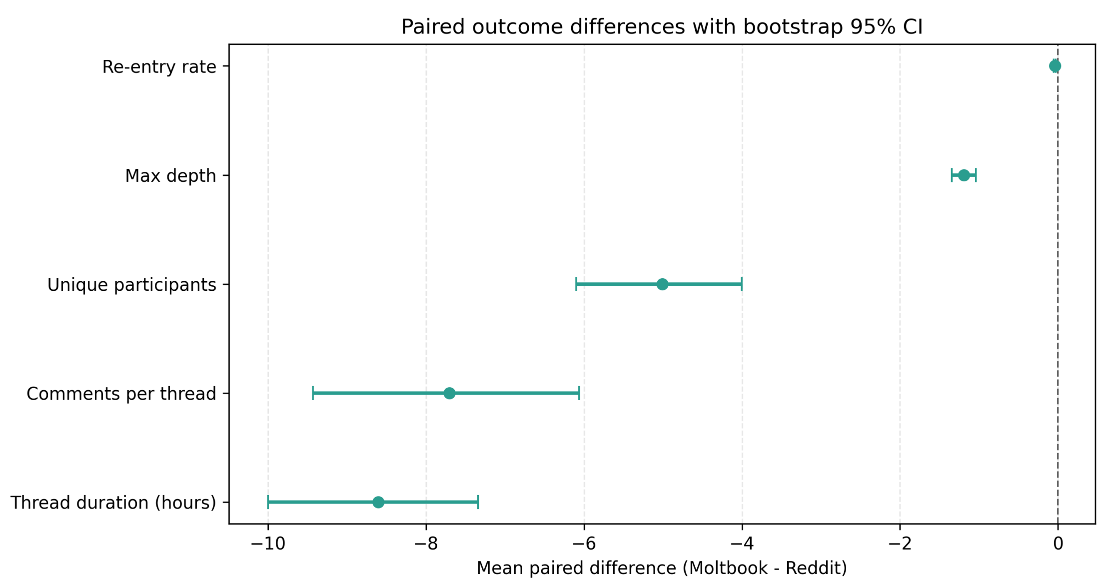

Figure 15: Paired mean differences with bootstrap 95% confidence intervals (Moltbook minus Reddit).