

When Does a Post Go Viral on Hacker News?

A Causal Study of Ranking Dynamics

CS 598 Causal Methods for HCI

Jonghu Jeong

jonghuj2@illinois.edu

University of Illinois Urbana-Champaign

Urbana, IL, USA

Abstract

Hacker News (HN) is a social news platform where user-submitted posts are ranked and displayed to a large technology-oriented community. HN posts become viral under a ranking system that can create feedback between visibility and engagement. This project studies HN virality using time-resolved rank and upvote trajectories collected from April 2026. We define viral posts as those reaching the Top-10 within 72 hours and analyze three stages of the virality process: trajectory regimes, exposure feedback, and submission timing.

First, we classify Top-500 entrant posts into organic, boosted, penalized, and uncertain trajectories using early rank-vote consistency features. The confidently classified groups show sharply different downstream outcomes, motivating a restriction to plausibly organic posts for downstream analyses. Second, we estimate whether near-boundary first-page exposure is associated with higher engagement. Among organic posts observed in both ranks 31–40 and ranks 21–30, the same post receives upvotes at an estimated 1.144 times the rate while in ranks 21–30. Third, under a specified causal DAG, we estimate that morning submissions have a substantially higher predicted probability of reaching the Top-30 within six hours than work-hours or after-work submissions.

Together, the results support a feedback-based view of HN virality. Viral outcomes are shaped not only by intrinsic post quality, but also by trajectory regimes, first-page exposure, and submission timing. The analysis also shows that causal interpretation differs across stages: trajectory classification acts as a filtering step, exposure feedback is studied through within-post comparison, and timing effects depend on the specified DAG.

1 Introduction

Hacker News (HN) is a high-traffic social news platform centered on technology, startups, and computing communities. External traffic estimates report 13.7 million visits to `news.ycombinator.com` in April 2026, and HN’s own moderator has described the site as receiving roughly 1,300 story submissions per day [1, 7]. In our April 2026 dataset (Section 3), we observe 30,088 submitted posts, confirming that posts compete in a dense stream for limited visibility.

HN organizes this stream through a ranking system. User submitted posts receive upvotes, age over time, and are displayed in ranked pages, with only the top 30 posts appearing on the main page by default. Because ranking depends on both votes and time decay, early engagement can determine whether a post reaches the main page before losing competitiveness [6, 11]. While it is natural to assume that posts become viral because they are preferred by

the HN community, the presence of a ranking system suggests an additional mechanism: posts that gain early visibility receive more attention, leading to more upvotes, which can further improve their rank and generate additional exposure. This creates a feedback loop in which early engagement, ranking position, and later engagement reinforce one another.

Motivated by this observation, we study how posts become viral on HN from the perspective of ranking dynamics. We define a post as *viral* if it reaches the Top-10 positions within 72 hours after submission. Rather than treating virality as a static outcome, we analyze the process through which posts enter high-visibility states and accumulate engagement over time. Our study is guided by three research questions:

- **RQ1 (Trajectory regimes).** What types of ranking trajectories do posts follow before becoming viral?
- **RQ2 (Exposure feedback).** Do posts receive upvotes at a higher rate while they are exposed near the Top-30 main-page boundary?
- **RQ3 (Submission timing).** Does submission timing affect the probability of first reaching the Top-30 within six hours?

To answer these questions, we collect time-resolved snapshots from the official HN API during April 2026 and reconstruct rank and upvote trajectories for submitted posts. RQ1 first separates organic, boosted, and penalized trajectories using early rank-vote consistency features. This step is important because moderator-driven boosting or suppression can contaminate downstream analyses of ranking dynamics. We then restrict RQ2 and RQ3 to the plausibly organic subset.

Our results support a feedback-based view of HN virality. RQ1 shows that posts follow heterogeneous trajectory regimes with sharply different downstream outcomes. RQ2 finds that, among organic posts observed on both sides of the near-boundary rank bands, the same post receives upvotes at a higher rate while in ranks 21–30 than while in ranks 31–40. RQ3 shows that morning submissions have a substantially higher predicted probability of reaching the Top-30 within six hours than work-hours or after-work submissions under the specified DAG.

Overall, this work makes three contributions. First, it reconstructs time-resolved HN rank and upvote trajectories from repeated API snapshots. Second, it identifies trajectory regimes that separate plausibly organic ranking dynamics from intervention-driven deviations. Third, it provides a causal framework for studying how exposure and submission timing shape viral outcomes, while making explicit where the causal interpretation is strongest and where it depends on modeling assumptions.

The rest of the paper is organized as follows. Section 2 reviews prior work on popularity dynamics, ranked systems, and Hacker News. Section 3 introduces the HN interface, ranking mechanism, data collection pipeline, and causal framing. Sections 4, 5, and 6 present the three empirical analyses. We then discuss the overall findings, causal interpretation, limitations, and implications before concluding.

2 Related Work

Popularity, Social Influence, and Feedback Loops. Prior work on social news shows that popularity is shaped not only by intrinsic quality but also by exposure and social influence. Stoddard analyzes Reddit and Hacker News and argues that observed popularity is an imperfect proxy for quality because platform dynamics affect attention allocation [8]. Muchnik et al. provide causal evidence that small early vote manipulations can produce persistent downstream differences in final outcomes [3], while Salganik et al. show that social influence increases inequality and unpredictability in collective outcomes [5]. Together, these studies motivate our view that HN virality reflects a feedback loop among early signals, rank, and later upvotes rather than content quality alone. Our RQ3 extends this feedback view to a pre-submission decision: whether timing affects the probability of entering the high-visibility state where feedback can begin.

Causal Inference in Ranked Systems. Prior work on ranked systems emphasizes that engagement differences across positions cannot be interpreted causally without accounting for exposure bias. Narayanan and Kalyanam use a regression discontinuity design to estimate position effects around ranking thresholds in search advertising [4], providing a useful template for local causal comparison near a cutoff. Joachims et al. address biased feedback in learning-to-rank by using propensity-based correction [2], highlighting that observed user responses are confounded by rank-dependent visibility. This literature informs our RQ2, where we compare within-post upvote rates across adjacent exposure bands near the Top-30 boundary rather than interpreting raw engagement differences across ranks as direct evidence of quality.

Hacker News-Specific Studies. Hacker News-specific work is relatively limited and consists mainly of empirical analysis, reverse engineering, and platform documentation. Stoddard provides the closest academic discussion of HN popularity dynamics [8], while Shirriff reverse engineers the ranking formula and argues that age, votes, and penalty-like adjustments jointly affect ranking outcomes [6]. The Hacker News FAQ also documents moderation and flagging behaviors that may alter visibility beyond the baseline ranking formula [11]. These sources support our hypothesis that posts may follow organic, boosted, and penalized trajectories, but they do not provide time-resolved trajectory reconstruction, trajectory-regime filtering, or a within-post analysis of how near-boundary first-page exposure relates to engagement rate.

3 Background

3.1 Interface and Interaction.

HN is a social news platform where users submit links or text posts (referred to as *posts*) and other users express interest by clicking

an *upvote* button. The primary interface consists of a *main page*, which displays the top-ranked posts, and a *new page*, which lists recently submitted posts in chronological order. Each page displays 30 posts, so the top 30 posts correspond to the first page of the main feed, while ranks 31–60 correspond to the second page. Similarly, the most recent 30 posts correspond to the first page of the new feed.

When users first visit HN, they are presented with the main page by default, while the new page must be accessed explicitly via navigation. As a result, visibility is highly concentrated on the main page.

3.2 Data Collection and Pipeline.

To support this analysis, we construct a dataset by collecting time-resolved snapshots from the official Hacker News API [10]. Specifically, we repeatedly query the main page and new page at fixed intervals (15 seconds) and retrieve metadata for each post, including submission time, upvotes, and other relevant attributes (April 1–30, 2026, full UTC days). Since HN does not provide direct measures of user exposure (e.g., view or impression counts), we use upvotes as a proxy for user attention throughout this work. Using these snapshots, we reconstruct rank trajectories and upvote dynamics over time, enabling us to analyze how posts move across ranks and accumulate engagement throughout their lifecycle. Table 1 summarizes how the April 2026 collection is filtered into the analysis samples used in RQ1–RQ3.

Cohort or sample	Count	Role
Submitted posts	30,088	Full April 2026 collection
Top-500 entrants	5,493	RQ1 trajectory classification cohort
Confidently classified posts	1,494	Organic, boosted, or penalized
Organic posts	946	Downstream RQ2/RQ3 analysis subset
RQ2 both-band posts	220	Within-post exposure-rate model
RQ2 near-miss posts	206	Descriptive baseline
RQ3 posts	946	Timing analysis sample

Table 1: Cohort flow from the April 2026 collection to the analysis samples.

3.3 Ranking Mechanism and Deviations.

HN ranking is commonly described by a score function that depends on both upvotes and post age, with a strong time-decay component [11]. A commonly referenced formulation is:

$$\text{score} = \frac{(v - 1)^\alpha}{(t + c)^\beta} \quad (1)$$

where v is the number of upvotes, t is the post age, and α, β, c are constants controlling the influence of votes and time. One commonly cited approximation uses $\alpha = 0.8$, $\beta = 1.8$, and $c = 2$ [6], although the exact formulation is not publicly specified.

This formulation implies that recency plays a dominant role: unless a post accumulates sufficient upvotes within a short time window (e.g., the first 24 hours), it becomes increasingly difficult to reach the top-ranked positions as it ages.

However, this baseline model does not fully explain observed rankings in practice. In addition to vote and time dynamics, HN

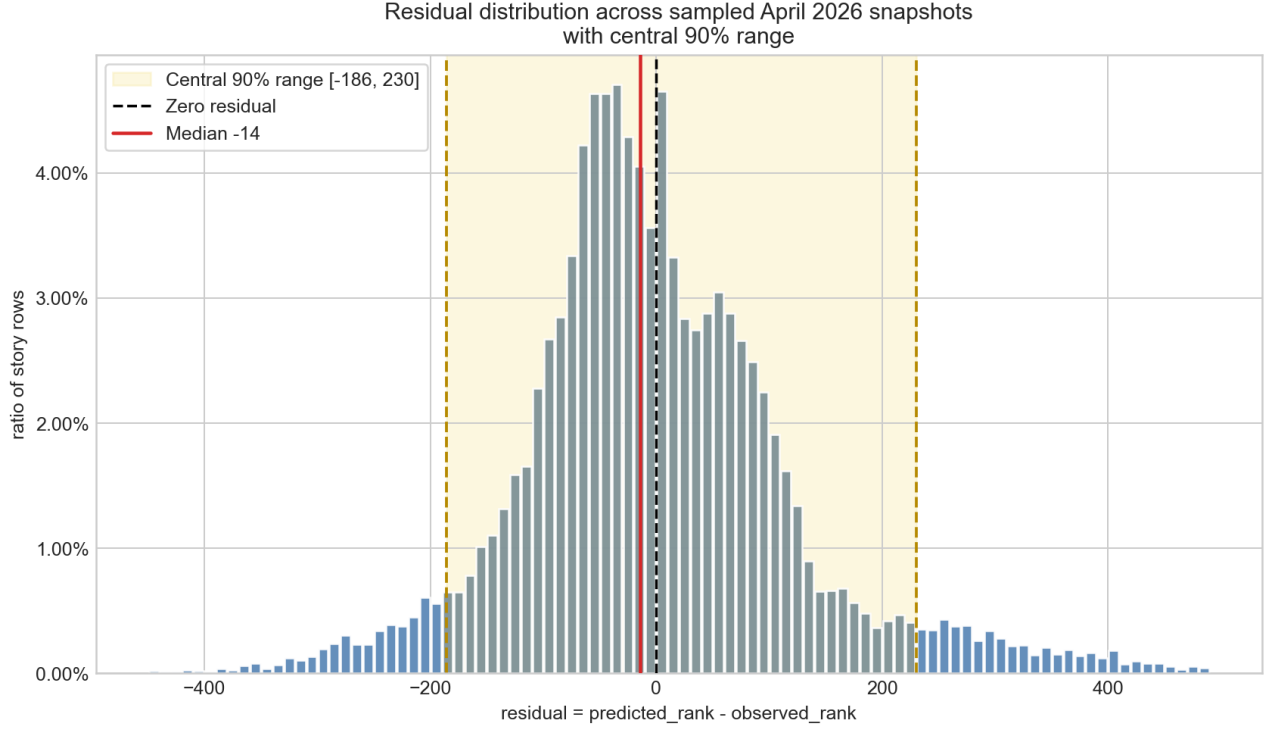


Figure 1: Residual between predicted rank (based on votes and age) and observed rank.

incorporates moderator-driven interventions that can significantly alter a post’s visibility. For example, moderators may promote posts through mechanisms such as the *second-chance pool*, where older or previously underexposed posts can be resurfaced near the top of the ranking (e.g., top 20) [9]. Conversely, posts may be penalized through flagging or other moderation signals, causing them to drop in rank despite accumulating upvotes [11].

Empirically, these effects manifest as large discrepancies between predicted and observed rankings. In our dataset, we compare ranks predicted from the baseline score function to actual observed ranks and examine their residuals (Figure 1). We find that many posts exhibit extreme deviations, such as posts predicted to rank near the bottom (e.g., around rank 400) appearing near the top (e.g., within top 20), and vice versa.

Motivated by these observations, we hypothesize that post trajectories can be broadly categorized into three regimes: *organic*, *boosted*, and *penalized*. Boosted and penalized posts are those affected by moderator interventions, while organic posts evolve according to the baseline dynamics of votes and time.

In this work, we focus primarily on organic trajectories, as they allow us to analyze ranking dynamics without the additional uncertainty introduced by latent moderator interventions. We defer the problem of identifying and separating these trajectory types to RQ1.

3.4 Scale and Attention Distribution.

HN operates at a meaningful scale, with approximately 13.7 million monthly visits [7]. Based on our dataset, we observe an average of 1,002.9 posts submitted per day, of which 5.04% reach the top 10 within 72 hours.

We use upvotes as a proxy for user attention and find that attention is highly concentrated at the top of the ranking. In particular, page 1 alone accounts for 67.74% of total upvotes, and the top five pages (top 150 posts) account for 96.27%. In contrast, posts ranked beyond the top 500 receive minimal attention, with a median of only 2 upvotes per item.

While these patterns may partly reflect content quality (i.e., better posts receive more upvotes), we observe a sharp decline in upvote rate when posts move from higher-ranked pages to lower-ranked pages, suggesting that rank-dependent exposure plays a significant role (we return to this confounding issue and address it more rigorously in RQ2). Furthermore, the new page is expected to receive substantially less traffic than the main page, reinforcing the role of ranking in shaping visibility. Overall, these results indicate that user attention is extremely concentrated in a small subset of top-ranked posts. All statistics in this paragraph are computed from our dataset unless otherwise specified.

3.5 Hypothesis on Viral Dynamics.

The observations above suggest that entry into the main page is the key bottleneck for virality on HN. Due to time decay, posts must accumulate sufficient early upvotes to reach high-visibility

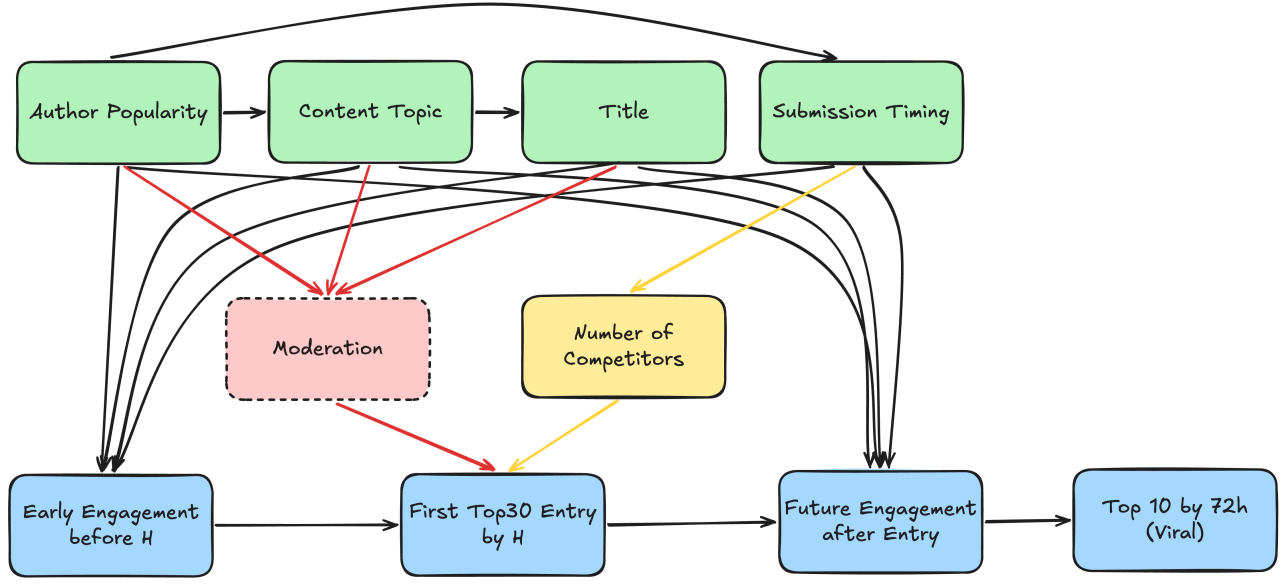


Figure 2: Causal DAG used to organize the assumed feedback process behind HN virality.

positions before losing competitiveness. Once on the main page, increased exposure can further accelerate upvote accumulation, creating a reinforcing feedback loop. Posts may reach the main page through different pathways, including organic growth driven by early upvotes and intervention-driven mechanisms such as boosting or suppression. Since such interventions are not directly observable and introduce additional uncertainty, we focus on understanding organic dynamics as the primary mechanism of interest.

We summarize this assumed process with a causal DAG (Figure 2): pre-submission factors such as author popularity, content topic, title, and submission timing affect early engagement; early engagement, latent moderation, and the competitive environment affect first Top-30 entry; and first Top-30 entry affects future engagement and eventual Top-10 viral outcomes. Based on this framing, our analysis proceeds in three steps: first, we characterize trajectory types and identify organic growth patterns (RQ1); second, we estimate whether near-boundary Top-30 exposure increases within-post upvote rate (RQ2); and third, we study whether submission timing affects first Top-30 entry within $H = 6$ hours (RQ3).

4 RQ1: What types of ranking trajectories do posts follow before becoming viral?

4.1 Setup

We analyze posts collected from April 1–30, 2026 (full UTC days), using time-resolved snapshots reconstructed from the Hacker News API. This yields 30,088 submitted posts in total, of which 5,493 appear in the Top-500 at least once and therefore enter the trajectory classification analysis.

Our goal in this section is not merely to describe trajectory patterns, but to identify a subset of posts whose dynamics are governed by the baseline ranking mechanism. Empirically, we observe substantial heterogeneity in ranking trajectories, including abrupt rank

jumps and drops that are not explained by vote accumulation or time decay (Figure 3).

This heterogeneity poses a challenge for subsequent analysis. In particular, boosted or penalized posts may receive exogenous changes in exposure due to moderator interventions, which would bias estimates of rank-based exposure effects (RQ2) and obscure actionable levers (RQ3). Therefore, we aim to identify and isolate *organic* trajectories, where rank, exposure, and upvotes evolve endogenously.

4.2 Trajectory Patterns and Rule Definition

Qualitative patterns. We begin by visualizing representative rank and upvote trajectories over time. Figure 3 shows three recurring patterns observed in the data:

- **Organic:** gradual rank improvement driven by steady upvote accumulation, followed by smooth decay.
- **Boosted:** sudden appearance at high ranks despite limited prior upvotes.
- **Penalized:** rapid drop in rank despite continued upvote accumulation.

These patterns suggest the presence of multiple underlying mechanisms beyond the baseline ranking formula.

Features. We construct early-stage features based on the first time a post appears in the Top-500 ranking. Let $first_seen_rank$ denote the rank at that time, and let $first_seen_age$ denote the corresponding post age in hours.

To quantify expected ranking strength, we use a score proxy based on a commonly referenced HN ranking formula [6]:

$$raw_proxy = \frac{(v - 1)^{0.8}}{(age_h + 2)^{1.8}},$$

where v is the upvote count and age_h is the post age in hours.

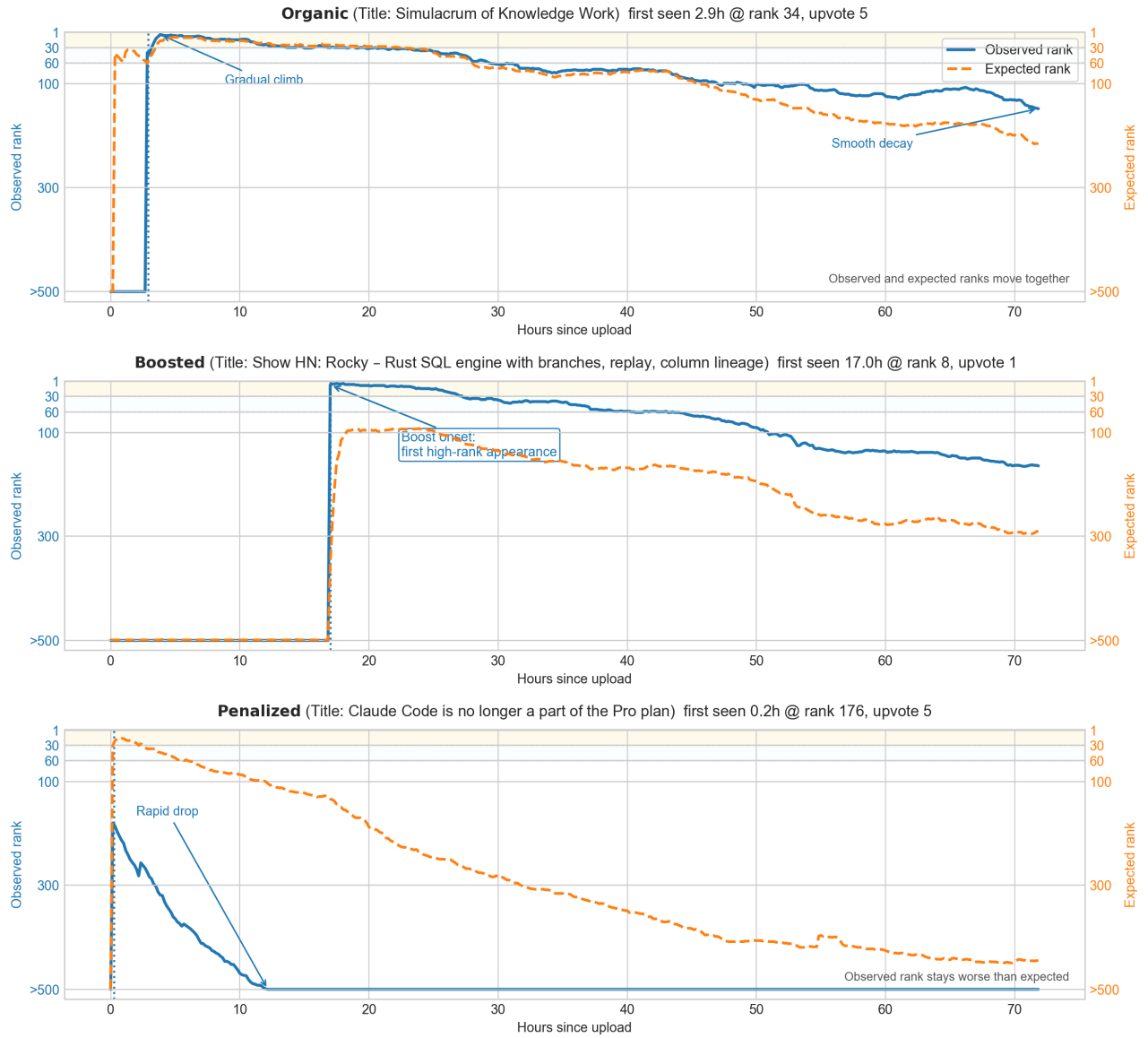


Figure 3: Representative rank and upvote trajectories for organic, boosted, and penalized posts.

At each snapshot, we compute $cutoff_k$ as the minimum proxy score among posts ranked within top- k , for $k \in \{30, 60\}$. We then define the entry ratio

$$entry_ratio_k = \frac{raw_proxy}{cutoff_k}.$$

We also define gap as the difference between the expected tier implied by the proxy score and the observed rank tier.

Final rule. Using these features, we define a rule:

- **Organic:** $gap = 0$ and $entry_ratio \in [0.67, 1.35]$
- **Boosted:** $first_seen_rank \leq 30$ and $entry_ratio_{30} < 0.20$
- **Penalized:** $first_seen_rank > 60$, $entry_ratio_{60} > 1.25$, and $first_seen_age \leq 6$ hours

- **Uncertain:** all remaining posts

This rule uses only early-stage signals and does not depend on final outcomes.

4.3 Validation of Classification

Coverage of trajectories. Applying the tuned rule to the Top-500 entrant cohort ($N = 5,493$), we obtain the following distribution:

- Organic: 17.2% ($n = 946$)
- Boosted: 4.2% ($n = 232$)
- Penalized: 5.8% ($n = 316$)
- Uncertain: 72.8% ($n = 3,999$)

Thus, the rule confidently classifies 27.2% of posts ($n = 1,494$).

Internal consistency. We evaluate post-label trajectory coherence using rank trajectories aligned at τ_i , where τ_i is defined as 60 minutes after the first time a post appears in the Top-500. We measure coherence using the within-to-between variance ratio of rank trajectories over the interval $[\tau_i, \tau_i + 24h]$, where lower values indicate more coherent within-group dynamics.

To assess whether the observed coherence is meaningful, we compare against a stratified-random null. Specifically, we generate 1,000 random label assignments that preserve the same label proportions within (*first_seen_rank*, *first_seen_age*) strata.

The rule achieves a within-to-between variance ratio of 1.296, compared to a stratified-random mean of 1.473 (10th–90th percentile: [1.375, 1.567]), corresponding to a 12.0% reduction relative to the null. The lower-tail p-value is 0.005, indicating that the observed grouping is more coherent than approximately 99.5% of stratified-random assignments. This suggests that posts assigned to the same trajectory type exhibit more similar post-label dynamics than expected by chance.

External validity. We examine downstream outcomes after τ_i . Post- τ_i Top-10 rates follow a consistent ordering:

- Boosted: 51.3%
- Organic: 30.0%
- Penalized: 0.9%

This ordering (boosted > organic > penalized) is stable across specifications.

Robustness. We evaluate robustness by applying a stricter variant of the classification rule that tightens thresholds in a consistent direction. Specifically, we adjust the tuned rule by tightening threshold parameters while keeping all other conditions unchanged:

- Organic: *entry_ratio* $\in [0.75, 1.15]$
- Boosted: *entry_ratio30* < 0.10
- Penalized: *entry_ratio60* > 1.45 and *first_seen_age* ≤ 4 hours

Under this stricter specification, group proportions change by at most 13.0 percentage points, and viral rates change by at most 4.5 percentage points. Importantly, the ordering of outcomes (boosted > organic > penalized) is preserved.

These results indicate that the qualitative conclusions are stable under reasonable perturbations of the classification thresholds.

4.4 Summary Statistics by Type

Table 2 summarizes post- τ_i lifecycle statistics for the three confidently classified groups. Lifetime denotes the median number of hours that a post remains in the Top-500 after τ_i , best rank denotes the median of each post’s minimum observed rank after τ_i (lower is better), and viral rate denotes the fraction of posts that reach rank ≤ 10 at least once after τ_i . The remaining 72.8% of posts are labeled as uncertain and are omitted from the table for clarity.

These differences are consistent with the intended interpretation of the labels. Boosted posts achieve the strongest downstream visibility, with a median best rank of 10 and a viral rate of 51.3%, despite accounting for only 4.2% of posts. Organic posts remain visible for longer, with a median post- τ_i Top-500 lifetime of 67.9 hours, and exhibit intermediate downstream success. Penalized posts, by contrast, have the weakest subsequent outcomes: their

Type	Share	Lifetime (h)	Best Rank	Viral
Organic ($n = 946$)	17.2%	67.9	33	30.0%
Boosted ($n = 232$)	4.2%	20.3	10	51.3%
Penalized ($n = 316$)	5.8%	18.3	124	0.9%

Table 2: Post- τ_i lifecycle statistics by trajectory type. The uncertain group contains 3,999 posts.

median best rank is 124, their median post- τ_i lifetime is 18.3 hours, and only 0.9% of them ever reach the Top-10 after τ_i .

Taken together, these statistics indicate that boosted and penalized posts follow substantially different visibility trajectories from organic posts, in ways that are not well explained by the baseline ranking mechanism alone. We therefore restrict subsequent analyses in RQ2 and RQ3 to the organic subset.

4.5 Discussion

Our results suggest that HN ranking dynamics are governed by a mixture of organic ranking-driven growth and intervention-driven deviations.

Boosted and penalized trajectories exhibit large deviations from the baseline ranking dynamics, consistent with exogenous shocks such as moderator interventions. In contrast, organic posts are more consistent with endogenous feedback between rank, exposure, and upvotes.

Since boosted and penalized posts may introduce intervention-driven contamination into downstream causal analyses, we restrict RQ2 and RQ3 to the plausibly organic subset. This allows RQ2 to estimate whether organic posts receive upvotes at a higher rate while in ranks 21–30 than while in ranks 31–40, and RQ3 to study whether submission timing affects first Top-30 entry under cleaner ranking dynamics.

The organic subset is not intended to be representative of the full population of posts. Rather, it isolates a subset with plausibly clean early dynamics to reduce confounding from unobserved interventions. Additionally, moderator actions are not directly observable, and the classification is based on heuristic rules, which may introduce misclassification in borderline cases.

5 RQ2: Does Top-30 exposure increase upvote rate?

5.1 Setup

RQ2 asks whether HN posts receive upvotes at a higher rate when they are exposed on the main page. We define ranks 21–30 as the treated exposure band and ranks 31–40 as the control exposure band. Both bands are close to the Top-30 boundary, but only ranks 21–30 are displayed on the first page. We focus on this near-boundary region to avoid comparing Top-30 posts with much higher-ranked posts whose baseline popularity may be substantially different.

We use only posts classified as *organic* in RQ1. For each post, we measure the total time it spends in each exposure band and the total number of upvotes it receives during that time. The primary comparison is within post: we use posts that are observed in both ranks 31–40 and ranks 21–30. This design compares the same post’s upvote rate below versus inside the main page boundary.

The primary sample contains 220 organic posts observed in both bands. These posts spend 55.56 total hours in ranks 31–40 and 494.80 total hours in ranks 21–30. We also report a descriptive near-miss baseline using 206 organic posts that reach ranks 31–40 but never enter the Top-30. These near-miss posts are not used to estimate the primary coefficient, but they provide context for posts that approach the main page without crossing the boundary.

5.2 Model

Let Y_{ib} be the total upvote gain for post i in band b , and let h_{ib} be the total exposure duration in hours. The treatment indicator is $D_{ib} = 1$ for ranks 21–30 and $D_{ib} = 0$ for ranks 31–40. We fit a Bayesian Negative Binomial rate model:

$$Y_{ib} \sim \text{NegativeBinomial}(\mu_{ib}, \phi)$$

$$\log(\mu_{ib}) = \log(h_{ib}) + \alpha_i + \tau D_{ib}.$$

The offset $\log(h_{ib})$ converts the outcome into an upvote-rate comparison. The post-level intercept α_i captures stable post-specific demand. The coefficient τ estimates how the same post’s upvote rate changes when it is in ranks 21–30 rather than ranks 31–40. We do not include additional covariates in the primary model; the comparison is based on within-post exposure-rate differences between the two near-boundary bands.

We use weakly informative priors:

$$\alpha_i \sim \mathcal{N}(\alpha, \sigma_{\text{post}}), \quad \alpha \sim \mathcal{N}(0, 2), \quad \sigma_{\text{post}} \sim \text{HalfNormal}(1),$$

$$\tau \sim \mathcal{N}(0, 0.5), \quad \phi \sim \text{Exponential}(1).$$

The post-level intercept is implemented with a non-centered parameterization. The sampler converged well: the maximum $\hat{R}$ is 1.000, the minimum bulk effective sample size is 2,165, the minimum tail effective sample size is 3,785, and there are no divergences.

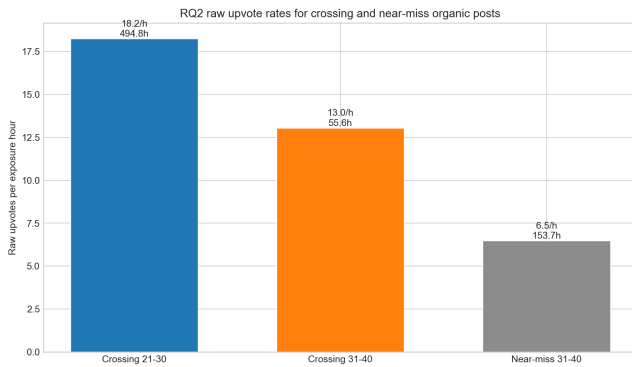


Figure 4: Raw upvote rates for organic crossing and near-miss posts. Crossing posts are observed in both ranks 31–40 and ranks 21–30. Near-miss posts reach ranks 31–40 but never enter the Top-30.

5.3 Results

Figure 4 shows the raw upvote rates by exposure group. Among crossing posts, the raw rate is 13.05 upvotes per exposure hour in ranks 31–40 and 18.25 upvotes per exposure hour in ranks 21–30. Near-miss posts have a lower descriptive rate of 6.49 upvotes per exposure hour in ranks 31–40. The lower near-miss rate suggests that posts which eventually cross into the Top-30 are already stronger than non-crossing near-miss posts, which is why near-miss posts are used only as descriptive context rather than in the primary coefficient.

Specification	τ	$\exp(\tau)$	$P(\tau > 0)$
21–30 vs. 31–40	0.133 [0.027, 0.237]	1.144 [1.022, 1.262]	0.977
26–30 vs. 31–35	−0.111 [−0.239, 0.031]	0.898 [0.771, 1.012]	0.095

Table 3: Bayesian exposure-rate estimates for RQ2.

Table 3 summarizes the posterior estimates. In the primary model, the posterior mean of τ is 0.133 with an 89% HDI of [0.027, 0.237], and $P(\tau > 0 \mid \text{data}) = 0.977$. On the rate-ratio scale, $\exp(\tau)$ has posterior mean 1.144 with an 89% HDI of [1.022, 1.262]. Thus, among organic posts observed in both bands, the same post receives upvotes at an estimated 1.14 times the rate while in ranks 21–30 compared with ranks 31–40.

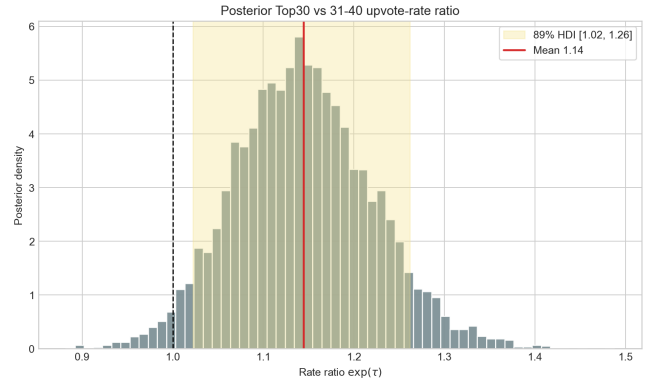


Figure 5: Posterior distribution of the Top-30 versus ranks 31–40 upvote-rate ratio. Values above 1 indicate a higher upvote rate while the same post is in ranks 21–30.

Figure 5 shows the posterior distribution of the rate ratio $\exp(\tau)$. Most posterior mass lies above 1, consistent with a higher upvote rate in ranks 21–30 under the primary specification.

As a robustness check, we repeat the same analysis using a narrower comparison: ranks 26–30 versus ranks 31–35. In this specification, the estimate is no longer positive: $\tau = -0.111$ with an 89% HDI of [−0.239, 0.031], and $\exp(\tau) = 0.898$ with an 89% HDI of [0.771, 1.012]. This suggests that the primary result reflects a broader exposure-rate difference between ranks 21–30 and ranks 31–40, rather than a sharp discontinuity exactly at the rank-30 boundary.

5.4 Discussion

The primary model estimates that organic posts receive upvotes at a higher rate while they are in ranks 21–30 than while they are in ranks 31–40. The estimated rate ratio is 1.144, suggesting that the same post receives about 14.4% more upvotes per exposure hour when it is on the first page near the Top-30 boundary. This result is consistent with an exposure feedback mechanism in which main-page visibility accelerates engagement.

This analysis is not a randomized experiment, but it is stronger than a simple cross-sectional comparison of high-ranked and low-ranked posts. First, we restrict the analysis to plausibly organic posts from RQ1, reducing contamination from intervention-driven rank changes. Second, the primary comparison is within post: each post is compared against itself across the 31–40 and 21–30 exposure bands. This absorbs stable post-specific demand, such as baseline content quality, author appeal, and title attractiveness. Third, both bands are close to the Top-30 boundary, so the comparison avoids contrasting first-page posts with much lower-ranked posts that may differ substantially in underlying popularity.

The near-miss comparison provides useful context. Posts that eventually cross into the Top-30 already receive more upvotes while in ranks 31–40 than posts that reach ranks 31–40 but never cross. This suggests that crossing posts are not exchangeable with near-miss posts; they are likely stronger even before entering the Top-30. For this reason, we use near-miss posts only as a descriptive baseline and estimate the primary effect from posts observed on both sides of the boundary.

At the same time, the robustness check limits the strength of the causal interpretation. When we narrow the comparison to ranks 26–30 versus ranks 31–35, the positive effect disappears. Therefore, RQ2 should not be interpreted as definitive evidence of a sharp discontinuous causal effect exactly at rank 30. A more cautious interpretation is that, among organic posts that move across the near-boundary region, first-page exposure is associated with a higher within-post upvote rate in the broader 21–30 versus 31–40 comparison.

This analysis has several limitations. First, the estimate applies only to plausibly organic posts observed in both exposure bands, not to all HN posts. Second, the organic subset is defined by a heuristic RQ1 classification and may still contain some misclassified trajectories. Third, upvotes are an engagement proxy rather than direct impressions, so we cannot directly observe exposure. Finally, although the within-post design reduces stable post-quality differences, it does not fully eliminate time-varying factors such as momentum, changing audience interest, or lifecycle stage. Thus, the result should be read as causal evidence consistent with exposure feedback, rather than as a fully identified platform-wide causal effect.

6 RQ3: Does submission timing affect first Top-30 entry?

6.1 Setup

RQ3 asks whether submission timing is an actionable factor for reaching the main page. We focus on first Top-30 entry within six hours after submission. This outcome captures whether a post

reaches the first page early enough to enter the exposure feedback process studied in RQ2.

We restrict the analysis to posts classified as *organic* in RQ1. The unit of analysis is a post. For each post, we define the binary outcome

$$Y_i = 1\{\text{post } i \text{ enters the Top-30 within 6 hours after submission}\}.$$

The analysis includes 946 organic posts submitted during April 2026. The observed Top-30-within-6h rate is 57.3%.

Submission timing is measured using Pacific Time. We model submission hour as a categorical variable with 24 hourly effects rather than as a linear numeric covariate. This avoids imposing an artificial linear trend over the 24-hour cycle. For interpretation, we summarize the hourly effects into three timing windows:

- **Morning:** 07:00–09:59 PT
- **Work-hours:** 10:00–16:59 PT
- **After-work:** 17:00–19:59 PT

Under the DAG in Figure 2, submission timing may affect first Top-30 entry through early engagement and the competitive environment. Because these are downstream mechanisms, we do not adjust for early engagement or number of competitors. The primary adjustment variable is author popularity, measured as z -scored $\log(1 + \text{karma})$.

6.2 Model

We fit a Bayesian hierarchical logistic regression:

$$Y_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) = \alpha + \theta_{\text{hour}[i]} + \beta \cdot \text{AuthorPopularity}_i.$$

The hourly effects are modeled hierarchically:

$$\theta_{\text{raw},h} \sim \mathcal{N}(0, 1),$$

$$\theta_h = \sigma_{\text{hour}} (\theta_{\text{raw},h} - \text{mean}(\theta_{\text{raw}})),$$

$$\sigma_{\text{hour}} \sim \text{HalfNormal}(1).$$

The remaining priors are

$$\alpha \sim \mathcal{N}(0, 2), \quad \beta \sim \mathcal{N}(0, 1).$$

The model estimates a DAG-dependent timing effect. We do not adjust for topic, title, early engagement, competitors, current rank, current upvotes, post age after submission, post type, domain, or day of week in the primary model.

We fit the model using four MCMC chains, 1,000 warmup draws per chain, and 2,000 posterior draws per chain. The sampler converged well: the maximum $\hat{R}$ is 1.000, the minimum bulk effective sample size is 2,501, the minimum tail effective sample size is 4,178, and there are no divergences.

6.3 Descriptive Timing Patterns

Figure 6 shows the observed Top-30-within-6h rate by submission hour. The hourly pattern is not flat: early morning hours tend to have higher observed success rates, while afternoon and early evening hours tend to be lower.

Aggregating hours into timing windows, the observed Top-30-within-6h rates are 69.8% for morning submissions, 46.3% for work-hours submissions, and 38.6% for after-work submissions. These

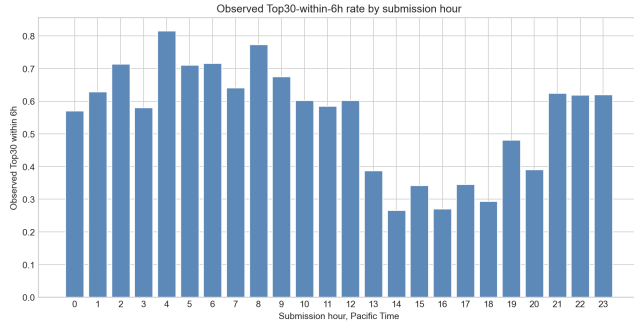


Figure 6: Observed Top-30-within-6h rate by submission hour in Pacific Time.

descriptive rates suggest a strong timing pattern before model adjustment.

6.4 Posterior Timing Effects

Figure 7 shows the posterior hourly effects on the log-odds scale. The model estimates positive hourly effects for much of the morning period and negative effects for afternoon and early evening hours.

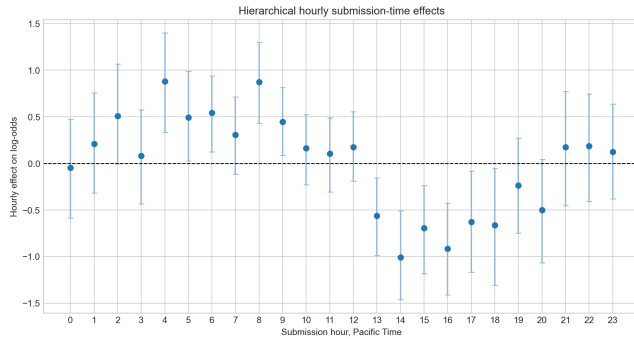


Figure 7: Hierarchical hourly submission-time effects on the log-odds of first Top-30 entry within six hours.

For the model parameters, the posterior mean of the author-popularity coefficient is negative: $\beta = -0.216$ with an 89% HDI of $[-0.329, -0.103]$. This means that, within the organic subset, higher author karma is associated with lower Top-30-within-6h probability after accounting for submission hour. This should not be interpreted as a general claim that author popularity hurts performance; it is a model-adjusted association within the selected organic subset.

6.5 Timing-Window Comparisons

We summarize the hourly posterior into three policy-relevant timing windows. For each posterior draw, we keep author popularity fixed and set all posts' submission hour to each timing window, then average the predicted Top-30-entry probability across posts and hours.

Figure 8 shows the posterior predicted probability of first Top-30 entry within six hours by timing window.

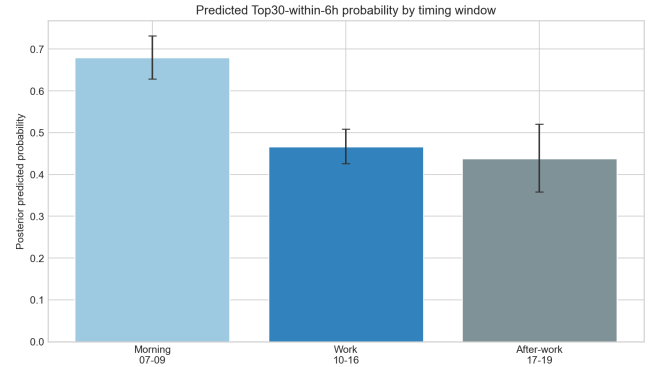


Figure 8: Posterior predicted Top-30-within-6h probability by submission timing window.

The posterior predicted probability is highest for the morning window. Morning submissions have an estimated Top-30-within-6h probability of 0.680 with an 89% HDI of $[0.629, 0.732]$. Work-hours submissions have an estimated probability of 0.466 with an 89% HDI of $[0.427, 0.509]$, and after-work submissions have an estimated probability of 0.437 with an 89% HDI of $[0.359, 0.520]$.

Timing window	Hours PT	Predicted probability
Morning	07–09	0.680 [0.629, 0.732]
Work-hours	10–16	0.466 [0.427, 0.509]
After-work	17–19	0.437 [0.359, 0.520]

Table 4: Posterior predicted probability of first Top-30 entry within six hours by timing window.

Table 5 and Figure 9 shows pairwise posterior risk differences between timing windows. Work-hours submissions are estimated to have a 21.3 percentage point lower Top-30-within-6h probability than morning submissions, with an 89% HDI of $[-0.279, -0.146]$. After-work submissions are estimated to have a 24.2 percentage point lower probability than morning submissions, with an 89% HDI of $[-0.345, -0.150]$. The difference between work-hours and after-work submissions is small and uncertain.

Contrast	Risk difference	$P(\text{RD} > 0)$
Work-hours – Morning	-0.213 [-0.279, -0.146]	0.000
After-work – Morning	-0.242 [-0.345, -0.150]	0.000
Work-hours – After-work	0.029 [-0.056, 0.112]	0.707

Table 5: Posterior risk differences between timing windows.

6.6 Discussion

The RQ3 results show a strong timing pattern among organic HN posts. Under the specified DAG and after adjusting for author popularity, morning submissions have a substantially higher predicted probability of entering the Top-30 within six hours than both work-hours and after-work submissions. The estimated difference is large: about 21 percentage points higher than work-hours and 24 percentage points higher than after-work.

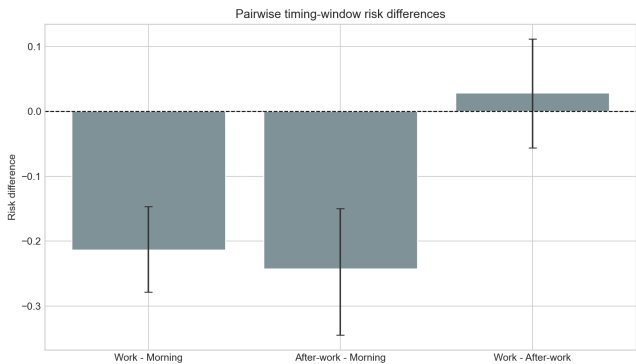


Figure 9: Pairwise posterior risk differences between submission timing windows. Negative values indicate that the first window has a lower predicted Top-30-within-6h probability than the second.

This pattern is actionable because submission timing is a controllable decision. Combined with RQ2, the result suggests a possible pathway for virality: posts submitted during favorable timing windows are more likely to enter the Top-30 early, and posts that enter the main page accumulate upvotes at a higher rate. In this sense, RQ3 connects pre-submission decisions to the exposure feedback process.

However, the causal interpretation is more assumption-dependent than in RQ2. Submission timing is not randomized. Authors may choose different times depending on topic, post type, intended audience, timezone, external promotion, or perceived newsworthiness. The primary model adjusts only for author popularity because the final DAG treats timing as having author popularity as its observed pre-treatment parent. If topic, title, author intent, or audience targeting also affect both timing and Top-30 entry, the timing estimates may still be confounded.

Therefore, RQ3 should be interpreted as a DAG-dependent timing analysis rather than a definitive causal estimate. The results are consistent with submission timing affecting first Top-30 entry, but they do not prove that changing the submission time alone would produce the same effect for every post. The estimates also apply only to the plausibly organic subset from RQ1 and should not be interpreted as full-population effects for all HN submissions.

7 Discussion

7.1 Main Findings

This project studies HN virality as a dynamic process shaped by ranking, exposure, engagement, and submission context. Across the three research questions, the results support the view that viral success is not only driven by intrinsic post quality, but also by whether a post enters high-visibility states early enough to benefit from ranking feedback.

RQ1 shows that HN trajectories are heterogeneous. Among 5,493 Top-500 entrants, our rule classifies 17.2% as organic, 4.2% as boosted, and 5.8% as penalized. These groups differ sharply in downstream outcomes: boosted posts have the highest Top-10 rate, organic posts show intermediate success, and penalized posts rarely

reach the Top-10. This motivates restricting RQ2 and RQ3 to the plausibly organic subset, where rank and upvotes are more likely to evolve through baseline dynamics rather than intervention-driven deviations.

RQ2 provides evidence consistent with exposure feedback. Among organic posts observed in both near-boundary exposure bands, the same post receives upvotes at an estimated 1.144 times the rate while in ranks 21–30 compared with ranks 31–40. This suggests that first-page visibility is associated with faster engagement accumulation. However, the effect disappears in the tighter 26–30 versus 31–35 comparison, so the result should be interpreted as evidence of a broader near-boundary exposure-rate increase rather than a sharp discontinuity exactly at rank 30.

RQ3 connects pre-submission timing to first Top-30 entry. Under the specified DAG and after adjusting for author popularity, morning submissions have the highest predicted probability of reaching the Top-30 within six hours: 0.680 for morning, compared with 0.466 for work-hours and 0.437 for after-work. This suggests that submission timing may be an actionable factor for entering the exposure feedback loop, although this interpretation depends strongly on the DAG assumptions.

7.2 Causal Interpretation

The causal strength differs across the analyses. Table 6 summarizes the estimand, identification strategy, and main limitation for each research question. We use this table to make explicit that the three analyses support the same feedback-based story, but differ in the strength and type of causal claim.

RQ	Estimand	Identification basis	Main limitation
RQ1	Trajectory regime classification	Early rank-vote consistency and downstream validation	Moderation is latent and labels are heuristic
RQ2	Within-post exposure-rate difference between ranks 21–30 and 31–40	Same-post comparison within plausibly organic posts	Time-varying momentum and lifecycle effects may remain
RQ3	Timing-window difference in first Top-30 entry probability	DAG-based adjustment for author popularity	Timing is self-selected and may be confounded by topic, intent, or audience targeting

Table 6: Summary of estimands, identification strategies, and main limitations across the three research questions.

RQ1 is not a causal effect estimate; it is a filtering step that reduces contamination from latent moderation or intervention-driven trajectories. RQ2 is stronger than a cross-sectional comparison because it compares the same post across nearby rank bands, which absorbs stable post-specific demand such as content quality, author appeal, and title attractiveness. Still, it cannot fully remove time-varying factors such as momentum, lifecycle stage, or changing audience interest.

RQ3 is the most assumption-dependent. Submission timing is not randomized, and authors may choose when to post based on topic,

post type, intended audience, timezone, or newsworthiness. Under the specified DAG, RQ3 estimates a timing effect after adjusting for author popularity, while treating early engagement and competition as mediators. If unobserved factors also affect both timing and Top-30 entry, the estimate may still be confounded. Thus, RQ3 should be read as a DAG-dependent timing analysis rather than a definitive causal estimate.

7.3 Implications and Limitations

The results suggest a plausible pathway for HN virality: submission timing affects the chance of early Top-30 entry, Top-30 entry increases exposure, and exposure increases engagement rate. For authors, this means that timing may matter because it changes the chance of entering a visibility feedback loop. For platform analysis, it means that final popularity should not be treated as a pure measure of community preference.

The study has several limitations. First, HN does not provide direct impressions or views, so we use upvotes as an engagement proxy. Second, the organic subset is defined by a heuristic rule and may still contain misclassified posts. Third, RQ2 applies only to organic posts observed on both sides of the near-boundary bands, not to all submissions. Fourth, RQ3 depends on the specified DAG and may be confounded by unobserved topic, author intent, or audience targeting. Finally, the analysis covers one month of data, so longer collection would be needed to test stability over time.

Future work could extend the dataset across multiple months, incorporate richer content and topic features, and model the rank-exposure-upvote feedback process dynamically. Direct exposure data, if available, would also allow a cleaner separation between visibility, attention, and preference.

8 Conclusion

This project analyzes how posts become viral on Hacker News using time-resolved rank and upvote trajectories from April 2026. The results support a feedback-based view of virality: posts become successful not only because they are high quality, but also because they enter high-visibility states early and benefit from additional engagement.

RQ1 identifies heterogeneous trajectory regimes and motivates focusing on plausibly organic posts. RQ2 shows that organic posts receive upvotes at a higher rate while they are in ranks 21–30 than while they are in ranks 31–40, although the evidence supports a broad near-boundary exposure-rate increase rather than a sharp cutoff effect. RQ3 shows that morning submissions have a substantially higher predicted probability of reaching the Top-30 within six hours than work-hours or after-work submissions under the specified DAG.

Overall, the study shows that HN virality is shaped by the interaction of pre-submission timing, early ranking dynamics, exposure, and engagement. It also illustrates why causal interpretation is difficult in ranked social systems: exposure, engagement, moderation, and timing are tightly coupled. By separating trajectory regimes, estimating within-post exposure-rate differences, and evaluating timing under an explicit DAG, this work provides a structured causal account of how ranking dynamics shape viral outcomes on HN.

References

- [1] dang. 2022. Comment on “Moderation is different from censorship”. Hacker News comment. <https://news.ycombinator.com/item?id=33454140> Describes Hacker News scale as roughly 5M monthly unique users, 10M page views per day, 1,300 submissions per day, and 13K comments per day. Accessed 2026-05-06.
- [2] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased Learning-to-Rank with Biased Feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 781–789.
- [3] Lev Muchnik, Sinan Aral, and Sean J. Taylor. 2013. Social Influence Bias: A Randomized Experiment. *Science* 341, 6146 (2013), 647–651.
- [4] Sridhar Narayanan and Kirithi Kalyanam. 2015. Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach. *Marketing Science* 34, 3 (2015), 388–407.
- [5] Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts. 2006. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science* 311, 5762 (2006), 854–856.
- [6] Ken Shirriff. 2013. How Hacker News Ranking Really Works: Scoring, Controversy, and Penalties. <https://www.righto.com/2013/11/how-hacker-news-ranking-really-works.html>. Accessed: 2026-04-05.
- [7] Similarweb. 2026. news.ycombinator.com Traffic Analytics, Ranking & Audience (February 2026). <https://www.similarweb.com/website/news.ycombinator.com/>. Estimated total visits: 12.9M; Accessed: 2026-04-05.
- [8] Greg Stoddard. 2015. Popularity and Quality in Social News Aggregators: A Study of Reddit and Hacker News. *arXiv preprint arXiv:1501.07860* (2015).
- [9] Y Combinator. 2021. Show HN: Second-Chance Pool. <https://news.ycombinator.com/item?id=26998308>. Accessed: 2026-04-05.
- [10] Y Combinator. 2024. Hacker News API. <https://github.com/HackerNews/API>. Accessed: 2026-04-05.
- [11] Y Combinator. 2026. Hacker News FAQ. <https://news.ycombinator.com/newsfaq.html>. Accessed: 2026-04-05.