Modeling Science Communication on Weibo: Causal Inference and Network Dynamics

Jingxin Yang
jingxinyang@stanford.edu
Stanford University — MS in Management Science & Engineering

Abstract

Weibo provides a large, semi-censored testbed for studying science communication in algorithmically mediated public spheres. Using a sentiment-labeled pilot sample (≈119,000 posts), we document descriptive patterns consistent with cold diffusion: posts containing technical keywords (e.g., *research*/*experiment*/*data*) are substantially longer, use fewer emoticons, and appear more often in deeper repost chains (median depth 1 vs. 0; maximum 9) than non-technical posts—suggesting that neutral/technical framing can travel widely even without overt affect. Building on this pilot, our preregistered design combines large-scale NLP, heterogeneous information networks, and quasi-experimental inference to test three mechanisms: (i) cold diffusion (does neutral/technical framing increase reach?); (ii) authority–influence decoupling (does platform amplification yield visibility without proportional organic engagement?); and (iii) grassroots amplification (do non-expert initiators trigger heavier-tailed cascades?). We operationalize algorithmic exposure via observable "trending" signals and a latent exposure score; identification leverages threshold-based recommendations, stacked event-study estimators for staggered timing, and retention weighting to adjust for post-hoc deletions. Outcomes span cascade depth, breadth and longevity (count and Hawkes models) and downstream scientist behavior (publications and collaborations). The study aims to clarify how platform governance and sociocultural context reshape the diffusion of scientific claims—and whether public attention measurably feeds back into research trajectories.

Keywords: Sina Weibo; science communication; information diffusion; causal inference; science of science

Core Mechanisms of Interest:

1. **Cold diffusion**: Emotionally neutral or technically framed content propagates more effectively under soft censorship.
2. **Authority–influence decoupling**: Expert posts gain algorithmic visibility but not proportional organic engagement.
3. **Grassroots amplification**: Non-expert actors can catalyse large cascades.
4. **Ideological silos**: Cascades stay within niche clusters with limited cross-cluster reach.

Operationalization of Mechanisms

Division of labor: Causal identification targets mechanism testing (e.g., treatment effects of tone, exposure, and account types) using fixed-effects and quasi-experimental designs. Predictive modeling (GCN/GAT over the heterogeneous network with text embeddings) is used to forecast cascade outcomes and to generate features fed back into identification as controls. We report causal estimates separately from predictive scores.

Definition: What counts as "science communication"?

We adopt a deliberately broad operational definition of "science communication" to match how scientific claims are used in public conversation. We code a post as science-related if it meets one or more of the following criteria:

1. Direct reference to scholarly outputs (paper titles, DOI links, CNKI/Scopus citations) or to preprints/technical reports; or
2. Media/press re-writes or popular-science adaptations of primary research (e.g., "" rewrites); or

3. Non-expert posts that invoke scientific evidence to support an argument (e.g., citing epidemiological statistics to justify policy claims); or

4. Use of graphical evidence (figures, charts, model outputs) as a stand-alone propagation unit; or

5. Instances of selective quoting or metaphorical use of science (e.g., using a research result as an analogy in political/moral argument).

We operationalize these via automated keyword/entity matching plus manual annotation on a stratified sample (Section Data).

Cold diffusion: We quantify sentiment with a Chinese BERT–based model that returns a polarity score $s_i \in [-1, 1]$. Posts are coded as neutral / technical when $|s_i| < 0.20$ **and** the text contains at least one domain keyword (, , , etc.); all others are labelled emotional. Formally,

$$\text{Cold}_i = \mathbf{1}\Big\{|s_i| < 0.20 \ \wedge \ \text{TechKeyword}_i = 1\Big\}.$$

The outcome is the log-transformed repost count $Y_i = \log\big(1 + \text{reposts}_i\big)$. Our baseline specification is

$$Y_i = \alpha + \beta\, \text{Cold}_i + \mathbf{X}_i\boldsymbol{\gamma} + \varepsilon_i,$$

where $\mathbf{X}_i$ controls for account type, follower count, posting hour, and topic fixed effects. A positive β would support the "cold diffusion" mechanism—that neutral/technical posts travel farther under soft-censorship conditions.

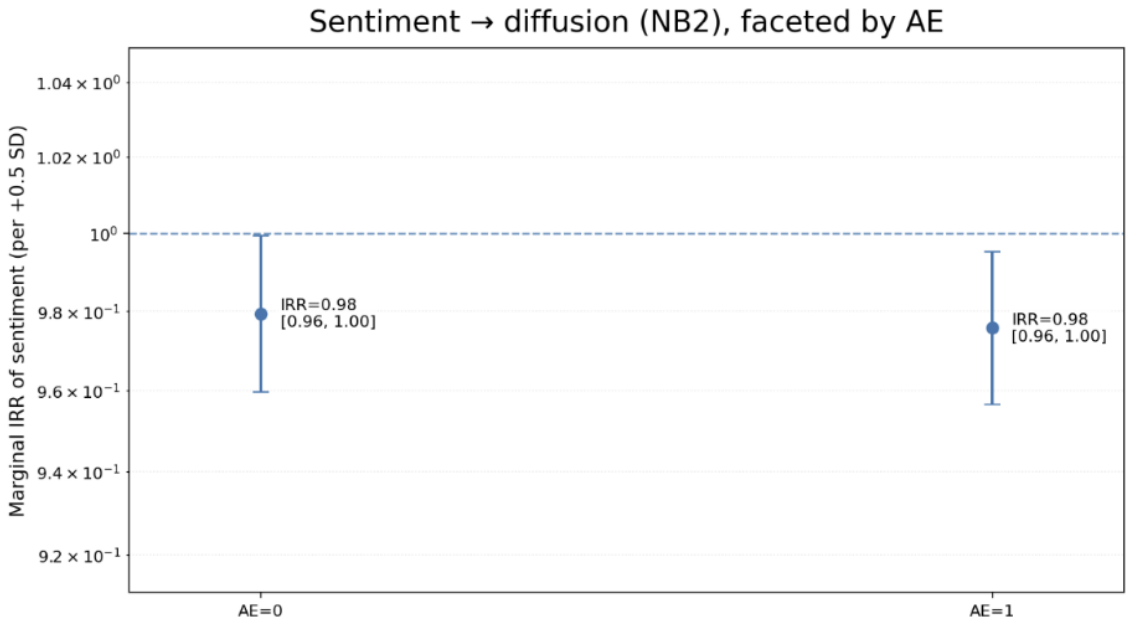


Figure 1: Sentiment strength → diffusion (NB2 marginal IRR), faceted by AE=0/1. Dots show the marginal incidence-rate ratio (IRR) for a $+0.5$ SD increase in the sentiment z-score; error bars are account-clustered robust 95% CIs; the dashed line marks IRR=1. Conditioning on account type, log followers, posting hour, and topic fixed effects, both panels yield IRRs near 1 (e.g., AE=0 $\approx 0.98\,[0.96, 1.00]$; AE=1 $\approx 0.98\,[0.96, 1.00]$), consistent with the *cold diffusion* mechanism.

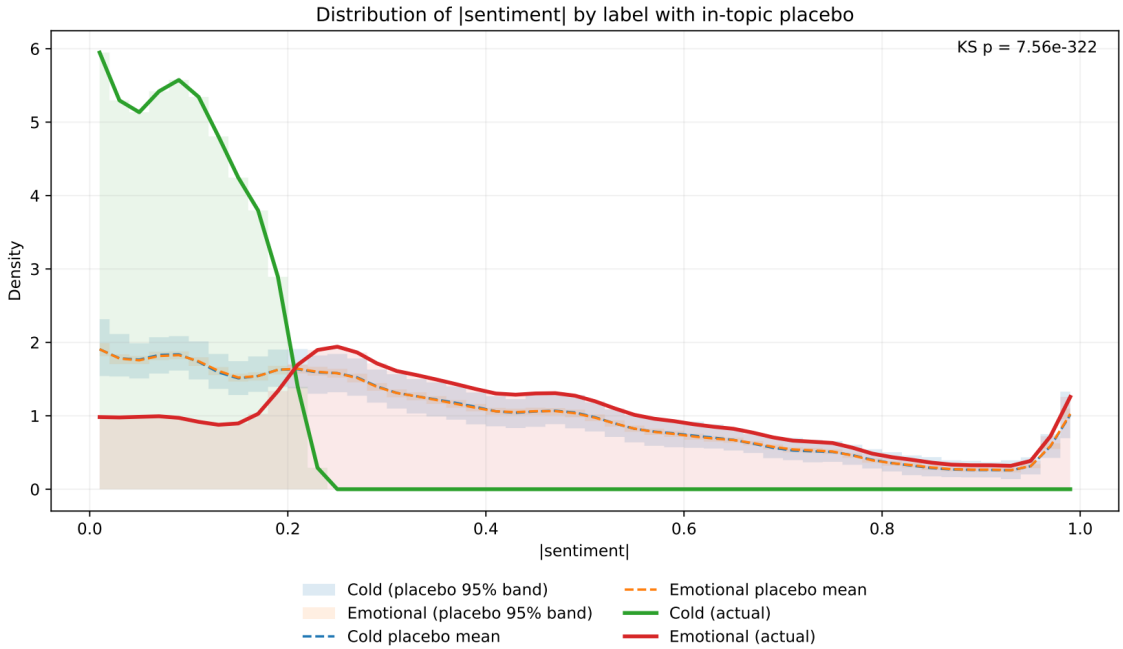


Figure 2: Distribution of $|s|$ (sentiment magnitude) by label (Cold vs Emotional) with an *in-topic* placebo. Solid curves are empirical densities; shaded bands are 95% permutation envelopes from 1,000 within-topic random relabelings of labels (dashed lines are placebo means). The gray vertical band marks the Cold rule ($|s| < 0.20$). Cold concentrates at low intensity, whereas Emotional extends to higher intensity and exceeds the placebo bands, supporting label validity beyond topic-composition artifacts. Two-sample KS (actual Cold vs Emotional): $p = 7.56 \times 10^{-322}$.

Authority–influence decoupling: Interaction term between algorithmic exposure and account type.

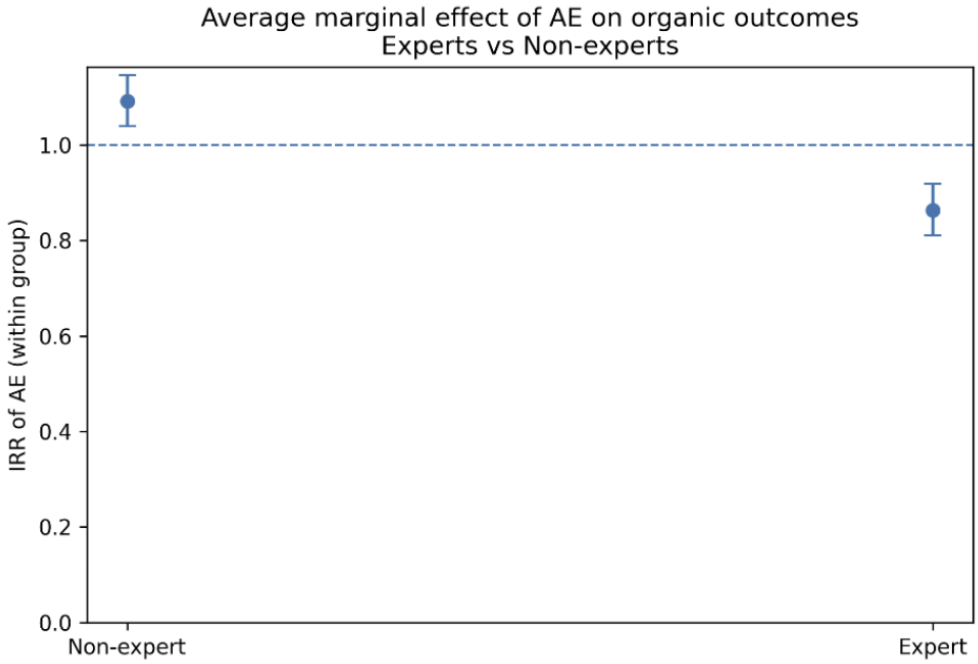
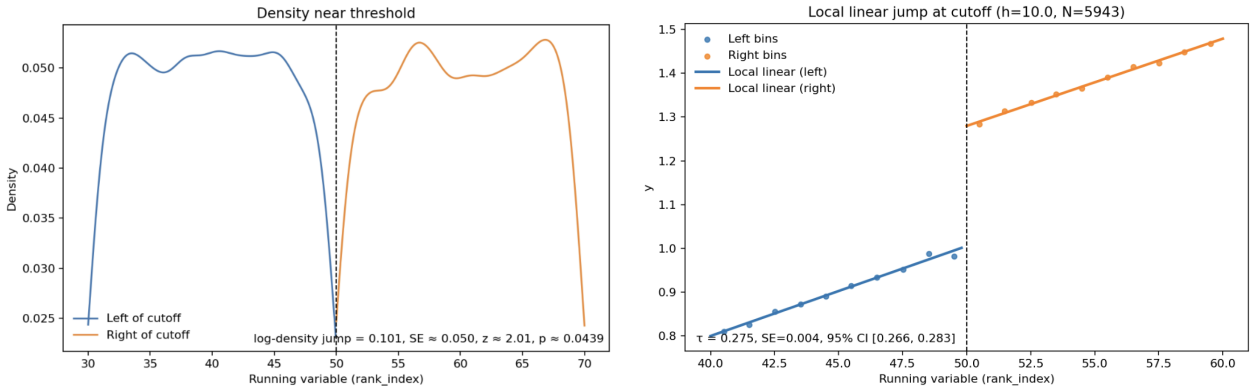


Figure 3: Average marginal effect (IRR) of algorithmic exposure (AE) on organic outcomes from a negative binomial model with an AE×Expert interaction. Within-group IRRs of AE: Non-experts $= 1.08\,[1.03, 1.13]$; Experts $= 0.86\,[0.82, 0.91]$. Controls: hour fixed effects, topic fixed effects, and log followers; N=12,000. Error bars are 95% Wald CIs; the dashed line marks IRR$= 1$.

Algorithmic exposure (AE) — planned measurement. We will operationalize AE using Weibo API fields (`is_hot`), parsed trending ranks (`rank_index` ≤ 50), and DOM badges (`icon_hot`). These fields are not present in the pilot corpus; acquiring them requires API access and/or real-time trending-page scraping.

Planned validation. We will validate AE by (i) manual spot checks on a stratified sample to confirm visible promotion, (ii) correlations with external proxies (short–term follower influx, media pickups), and (iii) sensitivity to alternative thresholds and a PCA-based continuous AE score.

AE validity checks (added). We validate AE indicators via three complementary checks: (i) manual inspection of a stratified sample (N≈1,000) to confirm that flagged posts visually display platform promotion; (ii) correlation analysis with external proxies (short-term follower influx, media pickup counts) to establish construct validity; and (iii) sensitivity analyses using alternative thresholds (top 5%, top 10%, median split) and the PCA-based continuous AE score. These steps help quantify measurement error and inform robustness ranges reported in Extended Data.



(a) Density near threshold

(b) Local linear jump

Figure 4: RDD around the trending cutoff (c=50). **Left:** McCrary-style density check using KDE on each side of the cutoff. Estimated log-density jump $\widehat{\Delta}_{\log f} = 0.101$ with SE$= 0.050$ (z=2.01, p=0.0439). **Right:** Local-linear RD for the outcome using bandwidth h=10: discontinuity $\hat{\tau} = 0.275$, SE$= 0.004$, 95% CI $[0.266, 0.283]$, effective N=5,943. Dashed vertical line marks the cutoff; points are binned means and lines are side-specific local linear fits.

Grassroots amplification: We label the initiating account as grassroots if it is neither a verified organisation nor an individual expert (i.e., `verified`=0 **and** no science-related tags in the profile). Define

$$\text{Grassroots}_i = \mathbf{1}\{\text{initiator } i \text{ is grassroots}\}\,.$$

Cascade size over the first 48 h is denoted $\text{Size}_{i,48h}$. Our main model is

$$\log\bigl(1 + \text{Size}_{i,48h}\bigr) = \alpha + \gamma\,\text{Grassroots}_i + \mathbf{X}_i\boldsymbol{\delta} + \varepsilon_i,$$

and we additionally compute the tail heaviness of the size distribution via the Pareto exponent $\hat{\kappa}$. A positive γ or smaller $\hat{\kappa}$ for grassroots initiators would confirm the mechanism that non-experts catalyse larger, heavier-tailed cascades.

Tail heaviness (Hill estimator). Let X denote cascade size; using the top-m order statistics,

$$\widehat{\kappa}^{-1} = \frac{1}{m}\sum_{j=1}^{m}\Bigl[\log X_{(n-j+1)} - \log X_{(n-m)}\Bigr], \tag{1}$$

where $X_{(k)}$ is the k-th order statistic and n is the sample size. We compare $\hat{\kappa}$ between grassroots and expert initiators (Fig. 5).

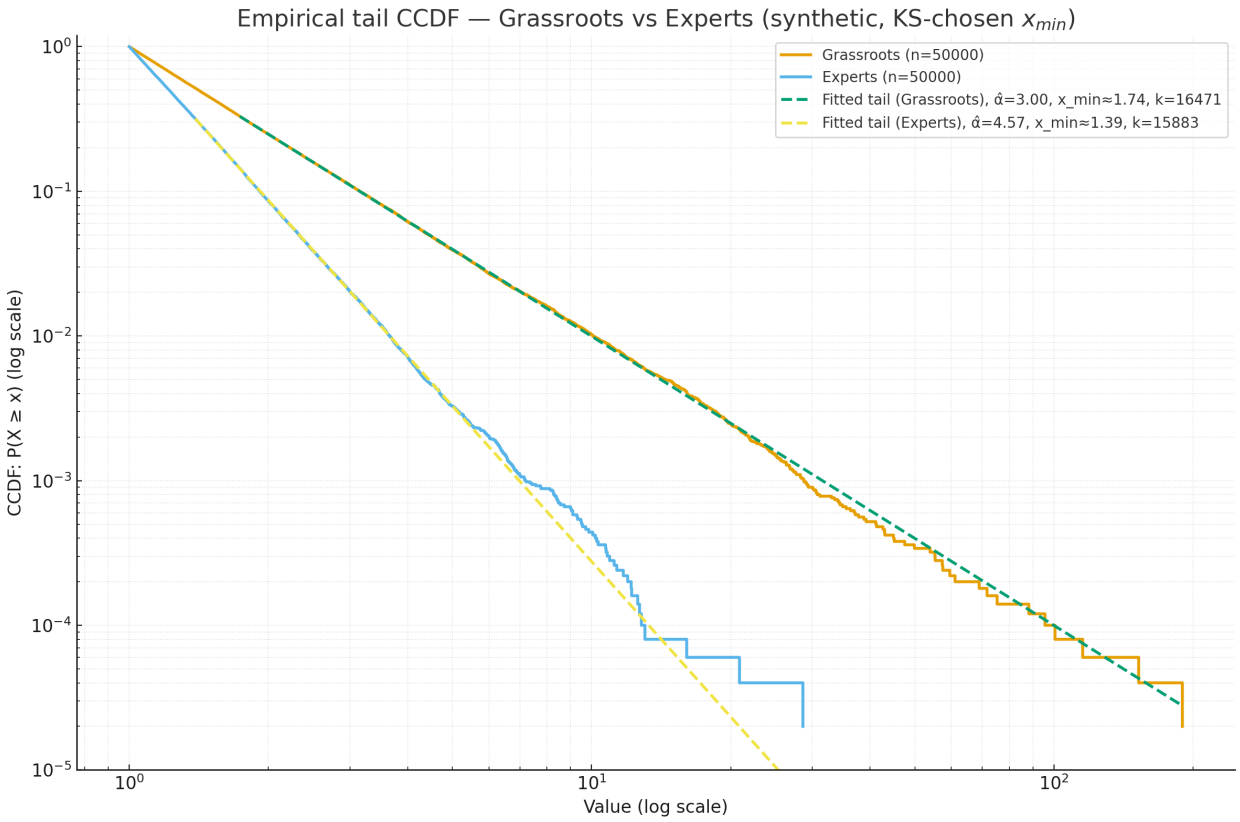
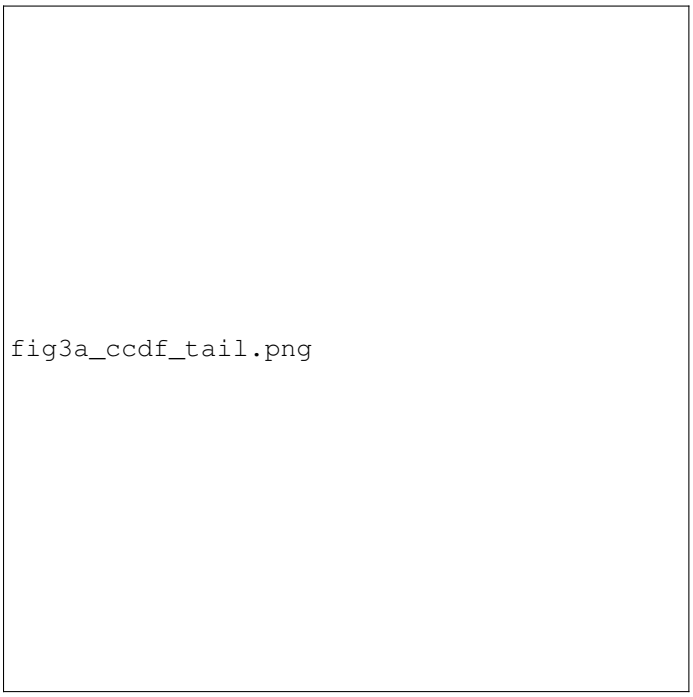


Figure 5: Empirical tail CCDF for grassroots vs. expert posts (log–log scale; $x_{\min}$ chosen by KS). Solid lines show empirical CCDFs; dashed lines are MLE power-law fits for $x \geq x_{\min}$ aligned at $x_{\min}$. The legend reports n, $\hat{\alpha}$, $x_{\min}$, and the tail sample size k. The grassroots tail is heavier ($\hat{\alpha} \approx 3.00$) than the experts ($\hat{\alpha} \approx 4.57$).



(a) Tail CCDF (log–log)

(b) Hill stability by initiator type

Figure 6: Tail heaviness and Hill stability for grassroots vs experts.

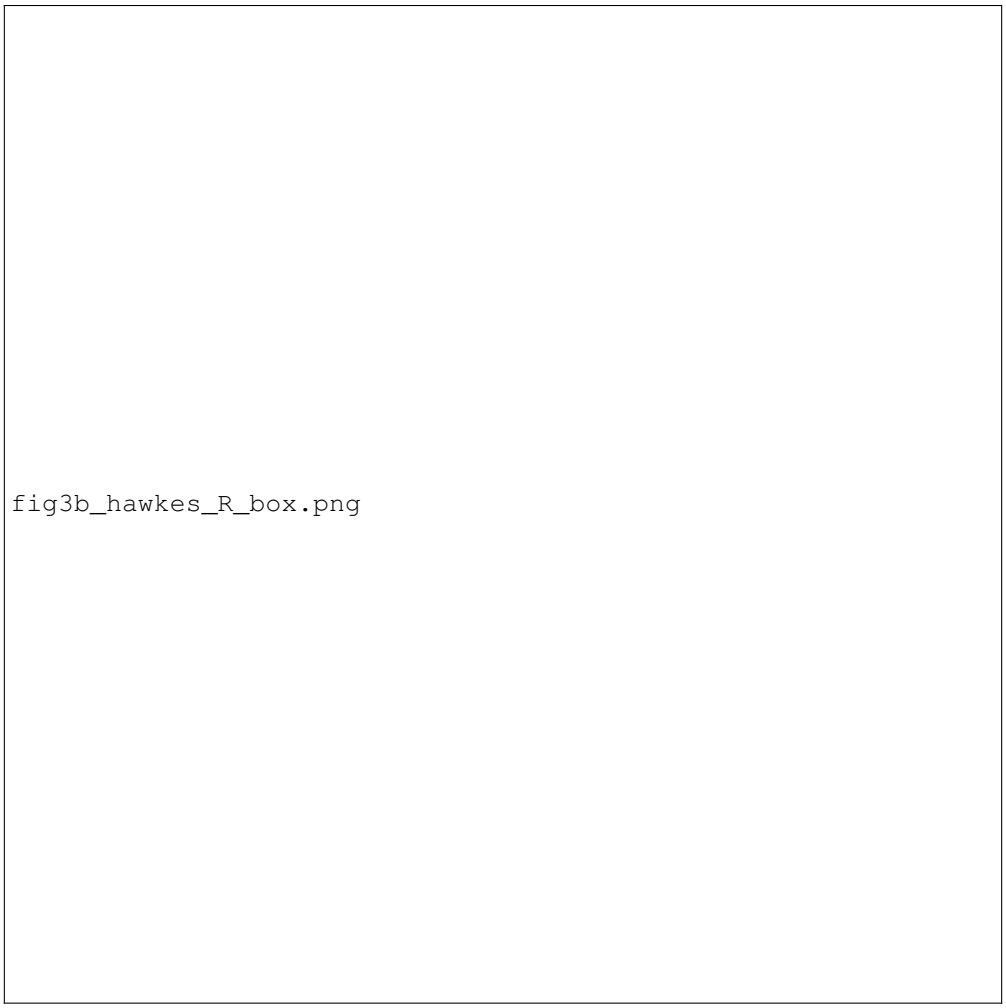


Figure 7: Hawkes reproduction number $\mathcal{R}$ by initiator type.

where $X_{(k)}$ is the k-th order statistic and n is the sample size. We compare $\hat{\kappa}$ between grassroots and expert initiators (Fig. 5).

Ideological silos: We run Louvain community detection on the follower graph (resolution = 1.0) and assign each user a community label $c(u)$. For every cascade we define N_i^{within} as the number of reposts from the initiator's own community and N_{ik}^{total} as the total from community k.

$$\text{CrossShare}_i = 1 - \frac{N_i^{\text{within}}}{\sum_k N_{ik}^{\text{total}}}.$$

We model

$$\text{CrossShare}_i = \alpha + \theta\, \text{TopicControls}_i + \mathbf{X}_i \boldsymbol{\lambda} + \varepsilon_i.$$

A low median CrossShare (e.g., < 0.25) indicates that diffusion is largely confined within ideological silos rather than bridging communities.

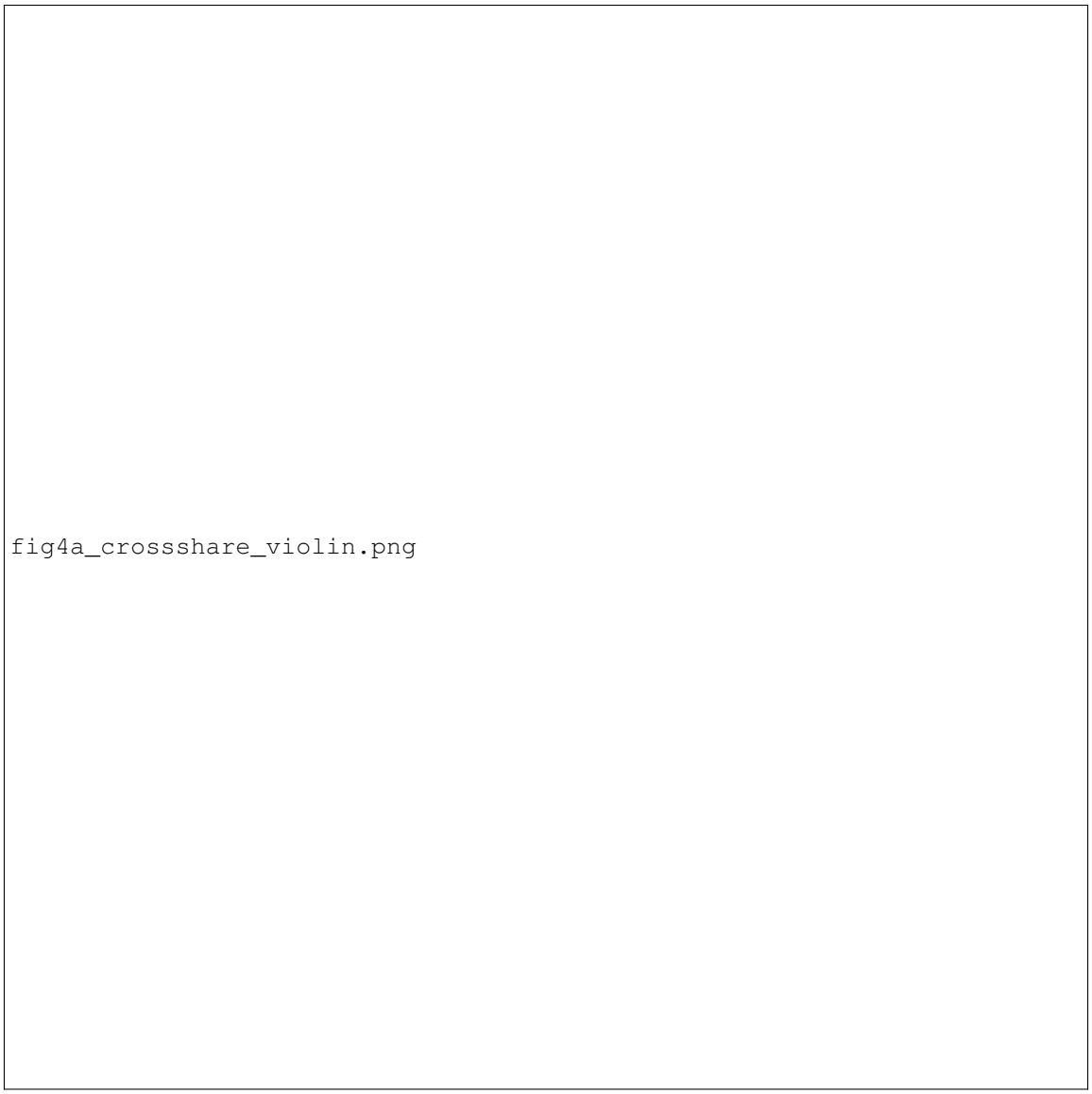


Figure 8: CrossShare distribution across events and initiator types.

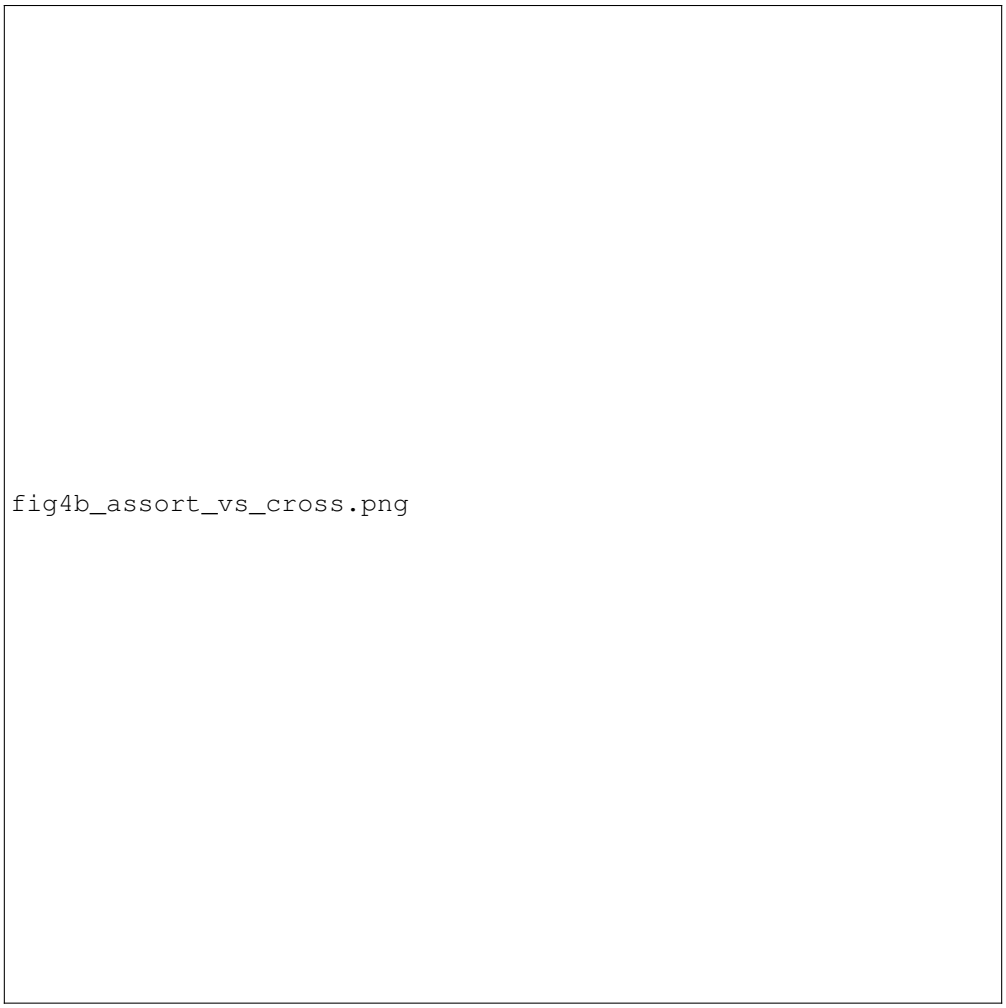


Figure 9: Event-level assortativity vs median CrossShare with permutation band.

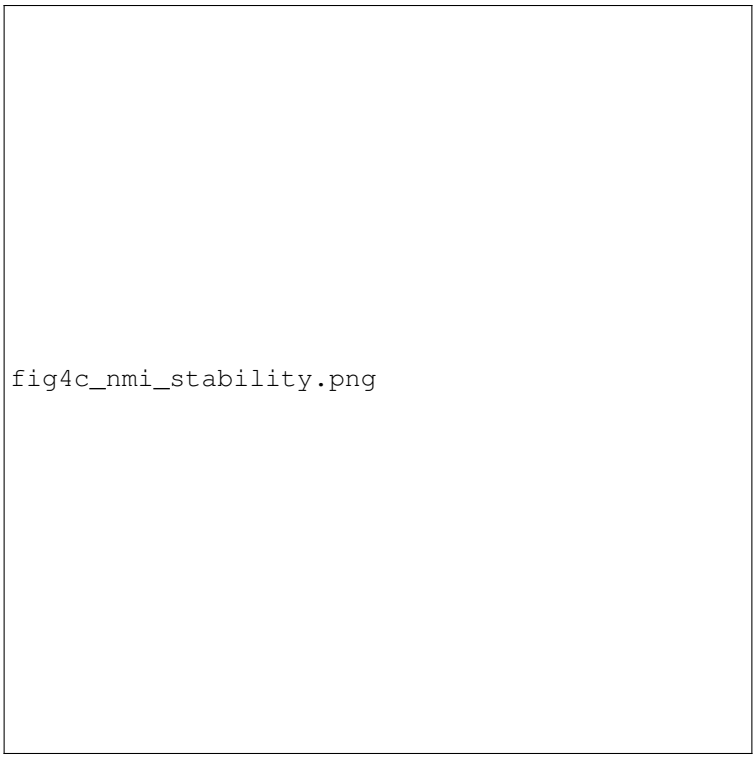


Figure 10: Community stability (NMI between $t-7$d and $t+7$d partitions).

Constructiveness and conversational phases. We operationalize "constructiveness" of an event's conversation as a composite index that captures (i) evidence salience (proportion of posts citing explicit data/reports), (ii) argumentative depth (average post length and presence of reasoned counterarguments), and (iii) dialogic exchange (fraction of posts with substantive replies vs. simple reactions). Formally, for cascade i,

$$\text{Constructive}_i \;=\; w_1 \cdot \text{EvidenceShare}_i + w_2 \cdot \text{ArgDepth}_i + w_3 \cdot \text{Dialogic}_i,$$

with the components defined as:

- EvidenceShare_i = fraction of posts in cascade i containing explicit references (paper title/DOI, institutional report, dataset link, or figure image with source);
- ArgDepth_i = mean log-length of substantive posts plus indicators for structured rebuttals (identified via discourse parsing);
- Dialogic_i = share of replies forming multi-turn chains (reply length ≥ 3) normalized by cascade size.

Each component (EvidenceShare_i, ArgDepth_i, Dialogic_i) is standardized (mean 0, sd 1) prior to weighting. We derive weights w_k from the first principal component estimated on the annotated pilot set and will report the PC loadings and the proportion of variance explained by the first PC in Extended Data. As robustness checks, we (i) compute an equal-weight composite and a median-based composite; (ii) fit an elastic-net model (5-fold cross-validation to choose penalty parameters) that predicts expert-annotated constructiveness and use its out-of-sample predictions as an alternative index; and (iii) report pairwise correlations (Spearman) among the alternative indices plus concordance with human labels (AUC / Spearman). All robustness results and the decision rule for selecting the primary index will be preregistered and documented in the Supplement.

We also hypothesise three temporal phases for attention dynamics (used in descriptive analysis and time-windowed models):

1. **Early spike (0–12 h)**: dominated by media coverage and mass reposting; expect high breadth, low deliberation.
2. **Community leader phase (12–36 h)**: community leaders and subject-matter actors (experts, NGOs, interest groups) enter, increasing evidence-salience and argument depth.
3. **Diffuse deliberation (36 h+)**: fractal, individual-level conversations and long-tail reposting; higher potential for niche retention but lower cross-cluster spread.

We will estimate phase-specific marginal effects (e.g., phase×Cold, phase×Expert) to test whether content features differently predict outcomes across stages.

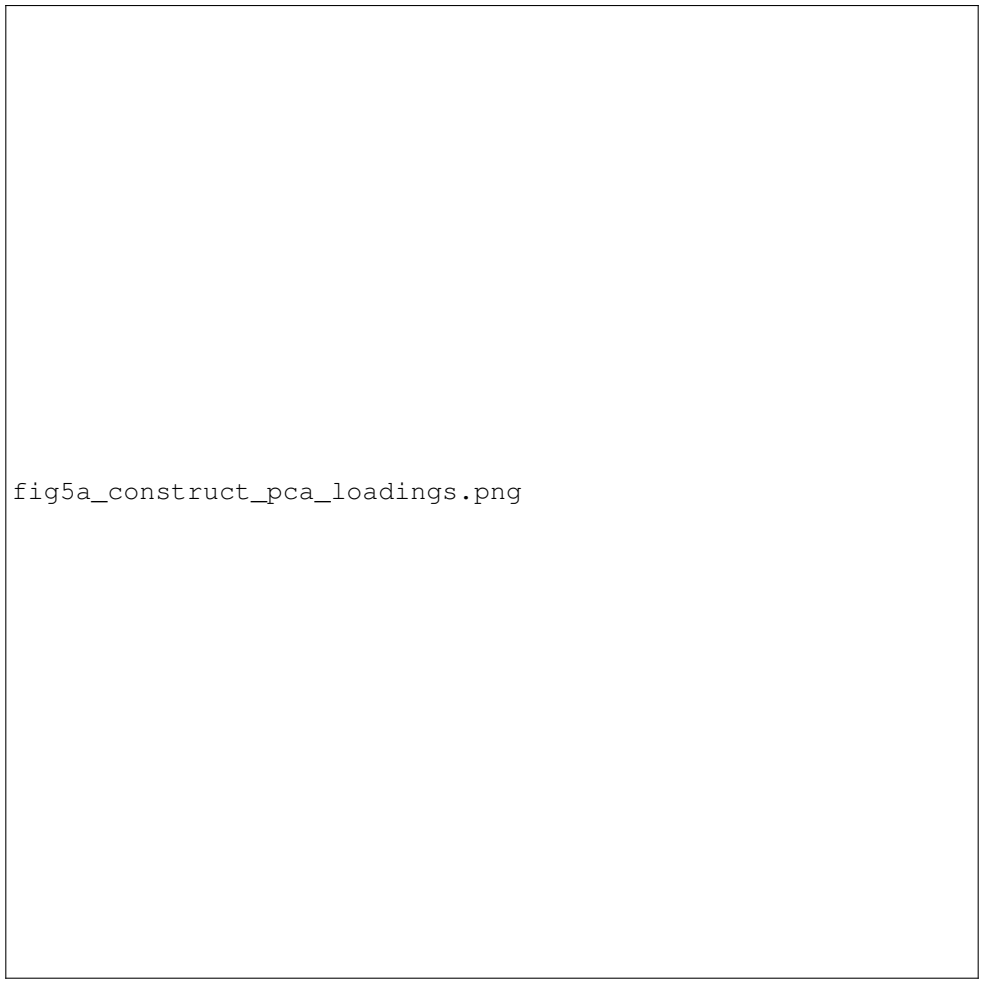


Figure 11: Constructiveness components: PCA loadings and variance explained (PC1).

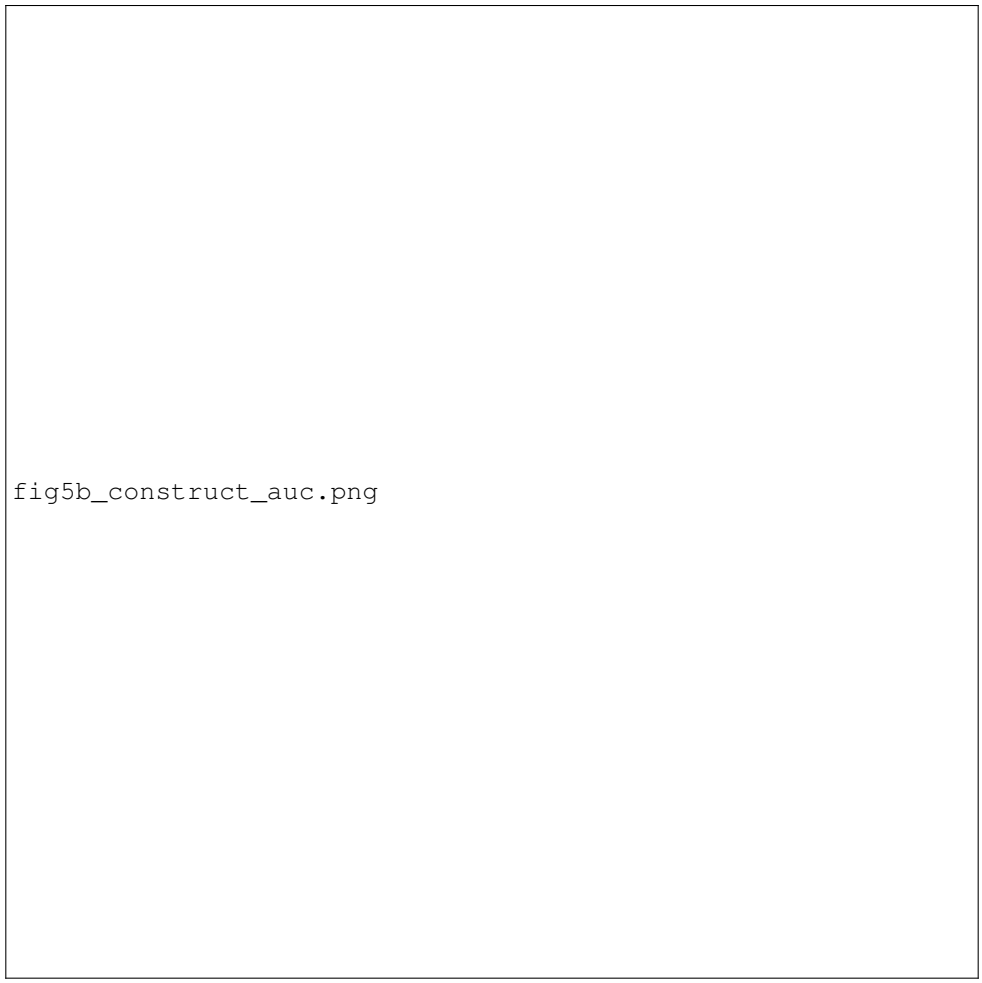


Figure 12: AUC comparison for constructiveness indices (PCA / Equal / Elastic-net).

Background and Motivation

Launched in 2009, Weibo is one of China's largest microblogging platforms, with over 516 million monthly active users as of late 2019. Its reposting mechanism facilitates large-scale public discourse and the rapid diffusion of information. Despite its societal influence, the platform's role in science communication remains underexplored in the empirical literature. Operating within a Chinese-language, state-regulated digital ecosystem, Weibo provides an important case for studying communication under distinct sociotechnical constraints.

In this semi-censored environment, content governance and recommendation systems jointly shape visibility, often privileging institutionally neutral or technically framed posts over affect-laden messages. As a result, Weibo constitutes a distinctive empirical site for examining how science circulates in controlled digital environments.

However, Western-centric virality theories—such as emotion-driven spread and authority-based influence—often assume liberal, minimally censored media systems. In China's semi-censored platforms like Weibo, where visibility is shaped by algorithmic amplification and political governance, these assumptions may not hold. The sociotechnical constraints of Chinese platforms, including soft censorship, identity verification, and opaque recommendation systems, fundamentally reshape who is seen, what circulates, and how users engage. This research therefore examines whether dominant diffusion theories—formulated in liberal, open-platform contexts—still hold under the constraints of tightly regulated digital environments like Weibo.

Such a perspective contributes to a more globally inclusive understanding of how scientific knowledge circulates

Table 1: Observable variables and coding scheme by mechanism

Mechanism	IV (coding)	DV	Controls	Data & tools
Cold diffusion	Cold_i: ($\lvert s_i \rvert < 0.20$ & tech keyword)	$\log(1 + \text{reposts})$	Account type, followers (log), posting hour, topic FE	Weibo API; BERT sentiment
Authority–Influence decoupling	AE_i, Expert_i, $\mathrm{AE} \times \mathrm{Expert}$	Cascade size / depth	Time, topic, followers	API flags + HTML parse
Grassroots amplification	$\mathrm{Grassroots}_i$ (initiator non-verified & no science tag)	$\log(1 + \mathrm{Size}_{48h})$; $\hat{\kappa}$	Same as above	API / crawler
Ideological silos	Initiator cluster $c(u)$ (Louvain)	$\mathrm{CrossShare}_i$	Topic, time, influence	Follower network; Louvain community detection
Constructiveness	$\mathrm{EvidenceShare}$, $\mathrm{ArgDepth}$, $\mathrm{Dialogic}$ (composite $\mathrm{Constructive}_i$ via PCA weights)	$\mathrm{Constructive}_i$	Phase (0–12h, 12–36h, 36h+), topic, account mix	Annotation + discourse parsing

across culturally diverse digital publics. Prior studies suggest that science communication on Weibo occurs through both event-driven mass communication and relationship-driven interpersonal diffusion, yet the causal mechanisms and structural patterns remain insufficiently understood. Leveraging large-scale data and advanced modeling, this study seeks to clarify these processes, offering insights into the flow of scientific knowledge in semi-censored online environments and their implications for platform governance and science policy in the Global South.

Beyond describing diffusion, we explicitly connect online "public use of science" to public funding and institutional attention. If platform-amplified, evidence-invoking conversations systematically precede shifts in funding signals (e.g., acknowledgments, topic alignment with new grants) and institutional uptake, this would demonstrate a feedback loop from public communication to the resource environment of science.

Literature Review

Research on science communication via Weibo has examined factors such as diffusion paths, user roles, and content sentiment. For example, Cui and Kertész (2021) analyzed Weibo activity during the Omicron outbreak and found that official media and scientific experts served as central nodes in information diffusion, with their posts shaping public emotional responses. Yu, Shen, and Wang (2017) studied how academic content is discussed and shared on Weibo and showed that Chinese users tend to engage with scientific topics that are entertaining, practical, or surprising. These studies underscore how content features and actor identities shape communication outcomes. Prior work has also shown that media consumption habits shape how the public engages with scientific information (Su et al. 2015).

However, prior work primarily relies on content analysis and descriptive social network metrics. There is a lack of causal inference to identify robust drivers of diffusion, and little use of heterogeneous information network (HIN) modeling that accounts for multiple node and relation types (e.g., users, posts, topics). As a result, these limitations hinder our understanding of the structural and semantic dynamics of science dissemination on Chinese platforms. Similar challenges of causal explanation in social media diffusion have been raised in studies on Western platforms. For instance, Vosoughi et al. (2018) found that false news on Twitter spreads more rapidly and broadly than true news, highlighting the role of novelty and emotional valence in information cascades (Vosoughi, Roy, and Aral 2018). Similarly, Zhao et al. (Z. Zhao, J. Zhao, Sano, et al. 2018) found that fake news exhibits distinct propagation patterns even at early stages of diffusion. Brossard and Scheufele (2013) further underscore how public understanding of science is shaped not just by message content but by the structure and affordances of digital media environments

(Brossard and Scheufele 2013). Our project seeks to address this gap by combining causal modeling and HIN-based analysis on large-scale Weibo data.

This methodological shift responds to recent trends in computational social science, where the focus has moved from descriptive metrics to explanatory and causal modeling, especially in identifying mechanisms that defy conventional assumptions.

Gaps and Opportunities: While prior studies have shed light on user roles and sentiment in information diffusion, they rarely test whether semi-censored platforms like Weibo exhibit surprising or inverted patterns of communication effectiveness. For instance, it remains unknown whether emotionally neutral or technically framed posts—rather than affectively charged ones—achieve greater reach in such contexts. Likewise, the possibility that grassroots communicators, rather than institutional experts, trigger larger cascades has not been rigorously examined. These untested mechanisms challenge dominant theories derived from Western platforms and create opportunities to discover alternative diffusion logics shaped by platform governance and sociocultural norms. Our study explicitly targets these blind spots through causal inference and heterogeneous network modeling.

Theoretical Contribution: Existing models of online virality emphasize emotional salience, novelty, and institutional authority as key drivers. However, such theories may inadequately capture dynamics in soft-censored ecosystems like Weibo. This study aims to evaluate and potentially revise several dominant assumptions:

- Emotion-centric virality: We test whether emotionally neutral or technically framed posts outperform emotional content in reach.
- Authority-based influence: We examine whether verified expert accounts retain influence in an environment where visibility is algorithmically assigned but not always trust-based.
- Open diffusion model: We investigate whether diffusion follows open cascade structures or is constrained within ideological silos.

These theoretical tests extend virality research into new governance and cultural contexts, providing globally comparative insight.

Research Questions

- What are the key drivers of scientific knowledge dissemination on Weibo?
- How do different source roles (e.g., experts, media, general users) influence dissemination outcomes?
- Do emotionally neutral posts diffuse more effectively in semi-censored platforms, contradicting emotion-driven diffusion models?
- Is algorithmic visibility sufficient to drive meaningful diffusion, or is authority decoupled from trust and engagement?
- Do grassroots communicators outperform institutional experts in catalyzing diffusion cascades?
- Do science posts propagate deeply within niche communities on Weibo but fail to reach broader audiences?
- What are the causal mechanisms linking content features (e.g., sentiment, narrative framing) and user interaction patterns to information spread?
- How do heterogeneous network structures affect the efficiency and depth of science diffusion?
- Can NLP, network models, and causal inference jointly explain or predict the effectiveness of science communication on social platforms?
- Do high-engagement science communication events on Weibo causally influence scientists' subsequent research trajectories, including topic choice, collaboration networks, and productivity?

- What pathways (e.g., media salience, funding shifts, institutional attention, collaboration opportunities) mediate the feedback from online experiences to science?
- Are feedback effects heterogeneous across career stage, discipline, gender, and baseline public visibility?
- Do surges in public "use" of science on Weibo translate into measurable changes in public funding attention and institutional uptake (e.g., acknowledgments, programmatic priorities)?

RQ–Hypothesis Mapping

Table 2: Research questions and corresponding hypotheses

Research Question (RQ)	**Hyp.**	**Testable implication / mechanism**
What roles do experts, media, and general users play in diffusion?	H1, H3	Account-type effects on cascade depth and breadth
Do emotionally neutral posts diffuse more widely?	H2	Neutral/technical tone may achieve wider reach under soft censorship
Is algorithmic visibility sufficient for engagement?	H5	Even with algorithmic boost, experts gain no extra organic engagement
Do grassroots actors trigger larger cascades?	H1 (counter-case)	Non-experts may spark larger cascades than experts
Are diffusion cascades trapped within ideological clusters?	H6	Measures cross-cluster vs. within-cluster spread (*CrossShare*)
Do viral communication events influence scientific behavior?	H4	Traces feedback loop from Weibo attention to publication behaviour (heterogeneity by event type: epidemiological vs. labor/policy)

Hypotheses

- **H1 (Source role).** Expert accounts are more likely to initiate diffusion cascades with greater depth than non-expert accounts; alternatively, grassroots initiators may compensate via network effects to produce comparable breadth.
- **H2 (Cold diffusion).** In semi-censored contexts, emotionally neutral or technically framed posts will on average be associated with broader dissemination than highly emotional posts.
- **H3 (Heterogeneous sentiment effect).** The effect of non-negative sentiment on diffusion breadth varies by source type (experts vs non-experts).
- **H4a (Feedback to science: topics/collaboration).** Experiencing a high-engagement event is associated with subsequent shifts in topic distributions, collaboration patterns, and productivity for affected scientists.
- **H4b (Feedback to science: funding/uptake).** High-engagement events that feature evidence-invoking posts precede increases in institutional attention and public-funding uptake signals (e.g., topical alignment with new grants, acknowledgments), relative to matched controls.
- **H5 (Authority–Influence Decoupling).** For platform-amplified posts, the marginal gain in organic engagement may not systematically differ between verified experts and ordinary users.
- **H6 (Ideological Silos).** Many science-related cascades exhibit low cross-cluster share; median CrossShare will be tested against permutation-based nulls.

Pilot Evidence from a Sentiment-Labeled Weibo Sample

We analyze a sentiment-labeled Weibo dataset (N $\approx$ 119,000 posts; roughly balanced between positive and negative sentiment). Sentiment labels correlate strongly with emoji usage (e.g., "[haha]", "[aini]" vs. "[lei]", "[nu]"), indicating that emotive symbols serve as explicit affective cues.

Cold-style content signals. Posts containing science or technical keywords (`yanjiu [research]`, `shiyan [experiment]`, `shuju [data]`) tend to be longer (median $\sim$104 vs. 54 characters) and contain fewer emoticons (median 1 vs. 2). They also appear more frequently in deeper repost chains (mean depth $\sim$1.5; median 1; maximum 9) than non-technical posts (mean $\sim$1.0; median 0), suggesting that neutral or technical content can propagate through larger cascades despite limited emotional charge.

Implication. Although descriptive rather than causal, these patterns are consistent with the *cold diffusion* hypothesis and motivate the planned causal tests that incorporate algorithmic exposure (AE), account attributes, and network structures.

Note: All Chinese lexical items are romanized in pinyin to indicate the original words appearing in Weibo posts.

Data Description and Collection Plan

We will collect publicly accessible Weibo data through a combination of official API access and custom-built crawlers. Our data sampling strategy includes two main channels: (1) posts containing predefined science-related keywords such as science, popular science, research, and (2) posts by verified or widely-followed science communicators, including institutional and individual accounts curated from existing lists and public directories. The time window will span from January 2015 to December 2025 to capture long-term trends and responses to major public science- and labor-related events (e.g., COVID-19 and vaccine debates; and high-attention labor/policy episodes such as the Jasic worker-organizing controversy (2018), the 996.ICU movement around tech-sector overtime norms (2019), the Foxconn labor controversy and associated suicides (circa 2010 and aftermath), and the Yue Yuen footwear workers' strike (circa 2014)). Data availability and post-hoc deletion rates will be pre-screened per event (see Supplement) and final event inclusion will be decided based on archival coverage and sample size.

Case / Event selection

For the main, cross-event analysis we will select a small set of "primary" events after pilot validation. Primary-event selection criteria: (i) post-pilot sample $\geq$5,000 after initial filtering; (ii) archival coverage $\geq$80% of raw crawl (or equivalent third-party archive coverage); and (iii) clear presence of science- or evidence-invoking posts (verified in manual inspection). Anticipated primary events for initial analysis (subject to pilot checks) are: COVID-19 (epidemiological), a CRISPR/gene-editing controversy (bioethical), and the Jasic worker-organizing controversy (labor/policy). Other candidate events (996.ICU, Foxconn, Yue Yuen) will be used for heterogeneity and robustness tests.

We select candidate events that (i) generated measurable, time-bounded spikes in public attention on Weibo; (ii) involved demonstrable references to scientific/technical evidence or policy arguments (e.g., occupational health data, epidemiology, labor statistics, technical reports); and (iii) cut across different sectors (manufacturing, electronics, tech services, and organized labor). Representative candidate events include the Jasic worker-organizing controversy (2018), the 996.ICU movement in tech (2019), the Foxconn labor controversy (circa 2010 and related aftermath), and the Yue Yuen workers' strike (circa 2014). We will pre-screen each candidate for data availability (volume, retention, and archival access) and report inclusion/exclusion decisions and keyword lists in Supplementary Materials. A Supplementary Table will list canonical Chinese keywords, hashtags, and initial pilot sizes for each event to ensure reproducibility.

To annotate whether a post constitutes "science communication," we will use a two-stage approach: automatic pre-filtering using keyword/topic matching, followed by manual annotation. A team of three bilingual annotators will independently label a stratified random subset (about 5,000 posts), following a coding guide developed in alignment with prior studies (Cui and Kertész 2021; Yu, Shen, and Wang 2017). Inter-rater reliability will be measured using Cohen's kappa, and disagreements will be resolved through consensus. This labeled set will be used both for validation and training classification models for full corpus labeling.

Pilot status. At the time of writing, pilot data collection is pending. We will run pilot crawls for three exemplar events

(one epidemiological: COVID-19; one bioethical: CRISPR/gene-editing; one labor: Jasic) to collect approximately 15,000–30,000 posts total. Pilot objectives: (i) confirm keyword coverage and scraper stability; (ii) estimate archival retention / deletion rates; (iii) produce an annotated sample (N≈5,000) for classifier training and constructiveness calibration. Pilot diagnostics (time-series plots, deletion-rate estimates, annotation guideline, and sample Cohen's) will be included in Supplementary Materials prior to submission.

Pilot feasibility and inclusion thresholds

Pilot feasibility checks (≈15,000–30,000 posts across three candidate events) will determine keyword lists, scraper configuration, and expected deletion rates. For an event to be included in the main cross-event analysis we require:

1. a minimum post sample of 5,000 after initial filtering, and
2. archival evidence sufficient to estimate deletion/retention rates (archived sample $\geq$ 80% of raw crawl or alternative third-party archive coverage).

Events failing the archival threshold will be analysed in archival-restricted robustness checks rather than in the pooled cross-event regressions.

Ethical data handling procedures will be enforced: only publicly visible posts will be used; personally identifiable information (PII) will be excluded or irreversibly hashed; deleted posts recovered via third-party archives will not be redistributed in raw form; and all user data will be anonymized before analysis. Data-sharing plans (aggregates, code, and instructions to reproduce analyses) will be deposited on OSF/GitHub; raw text will remain protected under approved data agreements where required.

When API access is constrained by rate limits or policy restrictions, we will employ web scraping tools (e.g., Scrapy, Selenium) to extract post data from public Weibo pages. We will also explore third-party datasets and archives where available (e.g., Weiboscope). All data will be anonymized and stored securely, in compliance with ethical guidelines.

In addition, we will identify Weibo accounts belonging to scientists or research groups (using verified information and profile keywords such as "", "", or institutional affiliations) and track their involvement in high-engagement events (top decile by reposts, comments, or likes). For these identified scientists, we will collect publication metadata from CNKI, Web of Science, Scopus, and Google Scholar between 2015–2025. This will allow us to examine whether online engagement correlates with subsequent shifts in research topics or collaborations.

Cross-platform comparison (exploratory)

Where feasible, we will conduct a cross-platform exploratory comparison using public Reddit data (for relevant English-language events) and other accessible platforms (e.g., public timelines on Douyin or Toutiao where possible) to check whether observed patterns are Weibo-specific or generalize across media ecologies.

Data acquisition tiers

1. **Primary:** Weibo Open API (rate-limited to ∼5,000 posts/day/account; targeting 2 million posts over 12 months).
2. **Secondary:** Custom scrapers using Scrapy/Selenium (estimated 100,000 posts/month, focused on high-engagement keywords).
3. **Tertiary:** Third-party datasets (e.g., Weiboscope, Weibo-COV), or historical dumps from collaboration partners.

Estimated total coverage ∼10 million Weibo posts (2015–2025), with metadata from ∼30,000 accounts, covering ∼200 major science- and policy-related events (explicitly sampling across epidemiological, bioethical and labor/policy episodes).

Scientist identification and matching. We compile a roster of scientists by cross-referencing institutional pages and public rosters, then resolve identities against Weibo profiles using multilingual name variants and affiliation strings. Ambiguities are flagged for manual adjudication. For each identified scientist, we build a longitudinal panel linking Weibo handles to bibliometrics via disambiguation heuristics (name+affiliation windows, coauthor overlap, and topic

similarity from embeddings). Matching controls include pre-event publication rate (past 3 years), field/discipline (journal classification), career age (years since first publication), baseline online visibility (mean monthly reposts prior to t_0), and institution tier. We expect to identify approximately 5,000–8,000 scientists with active Weibo presence, of which ∼2,000 will be linked to bibliometric records. We construct matched controls using nearest-neighbour propensity scores on pre-event trends (field, career stage proxies, baseline productivity, and prior online visibility).

Table 3: Main effects on diffusion outcomes (NB2 reported as IRR with 95% CI)

	NB2 (FE)	NB2 + Topic×Time FE	Quantile (0.75)
Cold_i	...	...	...
AE×Expert	...	...	–
AIC/BIC	...	...	–
Observations	...	...	...

Notes: Layout placeholder only. NB2 columns will report IRRs (95% CIs); Quantile column reports coefficients (not IRRs). Exact p-values and clustered SE settings will appear in Extended Data.

Methodology

Count model with overdispersion (NB2).

$$Y_i \sim \text{NB2}(\mu_i, \phi), \qquad \log \mu_i = \alpha + \beta \,\text{Cold}_i + \mathbf{X}_i\boldsymbol{\gamma} + \eta_{a(i)} + \tau_{t(i)}. \tag{2}$$

$$\text{Var}(Y_i) = \mu_i + \phi\,\mu_i^2. \tag{3}$$

Here, $\eta_{a(i)}$ are account fixed effects and $\tau_{t(i)}$ are time (e.g., month×year) fixed effects. $\mathbf{X}_i$ includes account type, follower count (log), posting hour, and topic fixed effects, unless otherwise noted. In the main text we report incidence rate ratios (IRR$= \exp(\beta)$) with 95% CIs and two-sided exact p-values. Standard errors are clustered at the account level; alternative clustering is reported in Extended Data.

Authority–influence interaction.

$$\log \mu_i = \beta_0 + \beta_1\,\text{AE}_i + \beta_2\,\text{Expert}_i + \beta_3\,(\text{AE}_i \times \text{Expert}_i) + \mathbf{X}_i\boldsymbol{\gamma} + \eta_{a(i)} + \tau_{t(i)}. \tag{4}$$

Cascade longevity (survival analysis)

We model time-to-last-repost with a Cox model:

$$h_i(t) = h_0(t)\exp\left(\theta_1\,\text{Cold}_i + \theta_2\,\text{Expert}_i + \mathbf{X}_i\boldsymbol{\theta}\right). \tag{5}$$

Cascade dynamics (self-exciting process)

We fit a Hawkes process to event times t_j:

$$\lambda(t) = \mu + \sum_{t_j<t} \alpha\, e^{-\beta(t-t_j)}, \qquad \mathcal{R} = \alpha/\beta. \tag{6}$$

We compare $\mathcal{R}$ across expert vs. grassroots and cold vs. emotional content.

NLP: We will use pretrained Chinese language models, such as Bidirectional Encoder Representations from Transformers (BERT) and LTP for segmentation, topic modeling, and frame detection. We will further analyze discursive frames (e.g., uncertainty, authority, controversy) to assess how different online framings may reshape public interpretation of science and subsequent engagement patterns. Sentiment and stance analysis will assess the emotional and attitudinal tone of content.

Network Analysis: We will construct a heterogeneous information network (HIN) to model the multi-relational structure of the Weibo ecosystem. The HIN will include three main types of nodes—users, posts, and scientific topics—and multiple edge types to capture user–post (e.g., publishes, reposts, comments), user–user (e.g., follows), and post–topic associations. This design allows the model to represent both social interactions and semantic affiliations in a unified graph structure. Figure 13 illustrates our HIN schema and the BFS-based diffusion layers used in subsequent analyses.

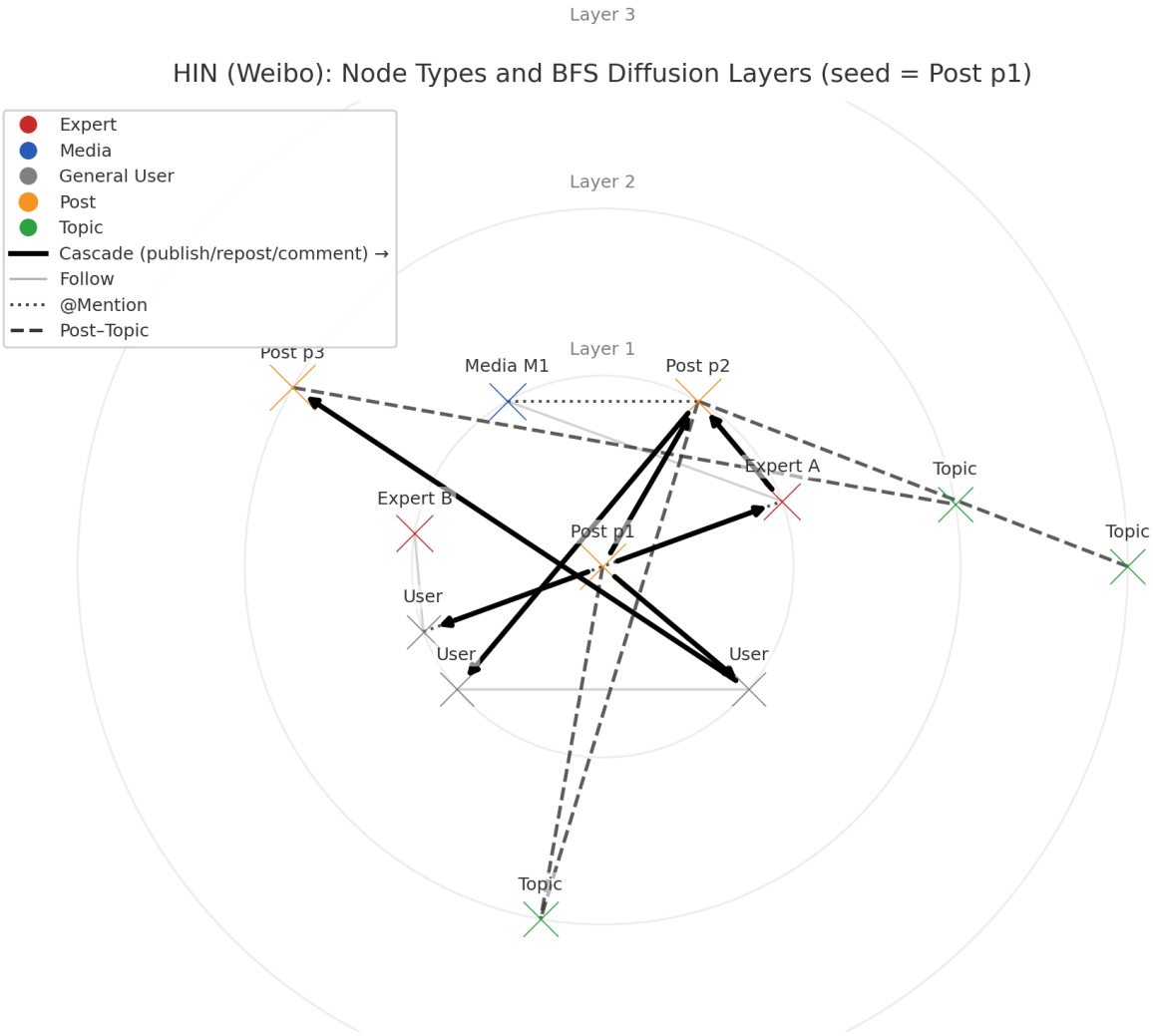


Figure 13: Heterogeneous Information Network (Weibo): node types and diffusion layers. Nodes: experts (red), media (blue), general users (gray), posts (orange), topics (green). Edge styles: thick solid with arrows = cascade edges (publish/repost/comment), light gray solid = follow edges, dotted = mentions, dashed = post–topic associations. Concentric rings denote BFS diffusion layers from the seed post (Layer 0).

To incorporate textual content into the network, we will use pretrained Chinese language models such as BERT to extract semantic embeddings from post text. ...

Causal Inference: We will employ causal inference frameworks (e.g., DoWhy) to construct directed acyclic graphs (DAGs) representing hypothesized causal structures among user attributes, content features, and diffusion outcomes. The design of these models will be grounded in established communication and diffusion theories. For example, based on Rogers' Diffusion of Innovations (Rogers 2003), we posit that users with higher scientific literacy or institutional affiliation may be perceived as more credible sources, increasing the likelihood that their posts are reposted. Similarly, agenda-setting theory suggests that exposure via platform amplification (e.g., trending lists or recommendations) may shape user attention and diffusion probability (McCombs 2005).

In our causal graph, treatment variables will include user-level attributes (e.g., verification status, follower count, content expertise), content features (e.g., sentiment, topic novelty), and platform-level amplification signals (if observable). Outcome variables will include repost count, cascade depth, and diffusion breadth. Control variables such as posting time, topic category, and baseline user activity will be included to reduce confounding bias. As a robustness check, we treat algorithmic exposure as a latent construct: we perform a one-factor model over `is_hot`, `rank_index`, and `icon_hot`, extract the first principal component, and re-estimate all specifications with this continuous AE score.

We will use DoWhy for causal effect estimation, implementing techniques such as propensity score matching and inverse probability weighting (IPW). Where possible, we will explore natural experiments—e.g., algorithmic interventions like Weibo's featured recommendations, trending cutoffs, or government-promoted content—as quasi-randomized treatments. Beyond propensity score matching and IPW, we will leverage placebo event tests, parallel-trend checks, and robustness analyses with topic–time fixed effects. To address network interference, we will employ neighbor-exposure intensity as an instrumental variable and validate findings with randomization tests on graph structures. Crucially, this framework links online exposure to subsequent offline scientific behaviors, thereby directly testing whether online experiences alter the trajectory of science itself.

Implementation note (Callaway–Sant'Anna). For staggered treatment/event timing, we will implement Callaway–Sant'Anna estimators using the 'did' R package and report group-time average treatment effects with event windows [-12,+24] months (and bootstrap CIs). Results from stacked and aggregated implementations will be contrasted and presented in Extended Data.

Survivorship / deletion adjustment (specific). To adjust for post-hoc deletions, we estimate retention probabilities $\hat{r}_{e,t}$ per event e and time bin t using comparisons between raw crawl snapshots and archived coverage (third-party archives). We then apply inverse-probability weighting (IPW) with weights $1/\hat{r}_{e,t}$ in weighted regressions to correct for differential retention; additionally, we present archival-only restricted estimates as lower-bound robustness checks. We will include the following diagnostic outputs in Supplementary Materials: (i) per-event time series of raw vs archived post counts; (ii) estimated retention probability matrices $\hat{r}_{e,t}$ (event × time-bin); (iii) deletion-rate heatmaps by account type and topic; and (iv) sensitivity tables comparing unweighted, IPW-weighted, and archival-only estimates. These diagnostics will demonstrate the feasibility and limits of deletion adjustments for each included event.

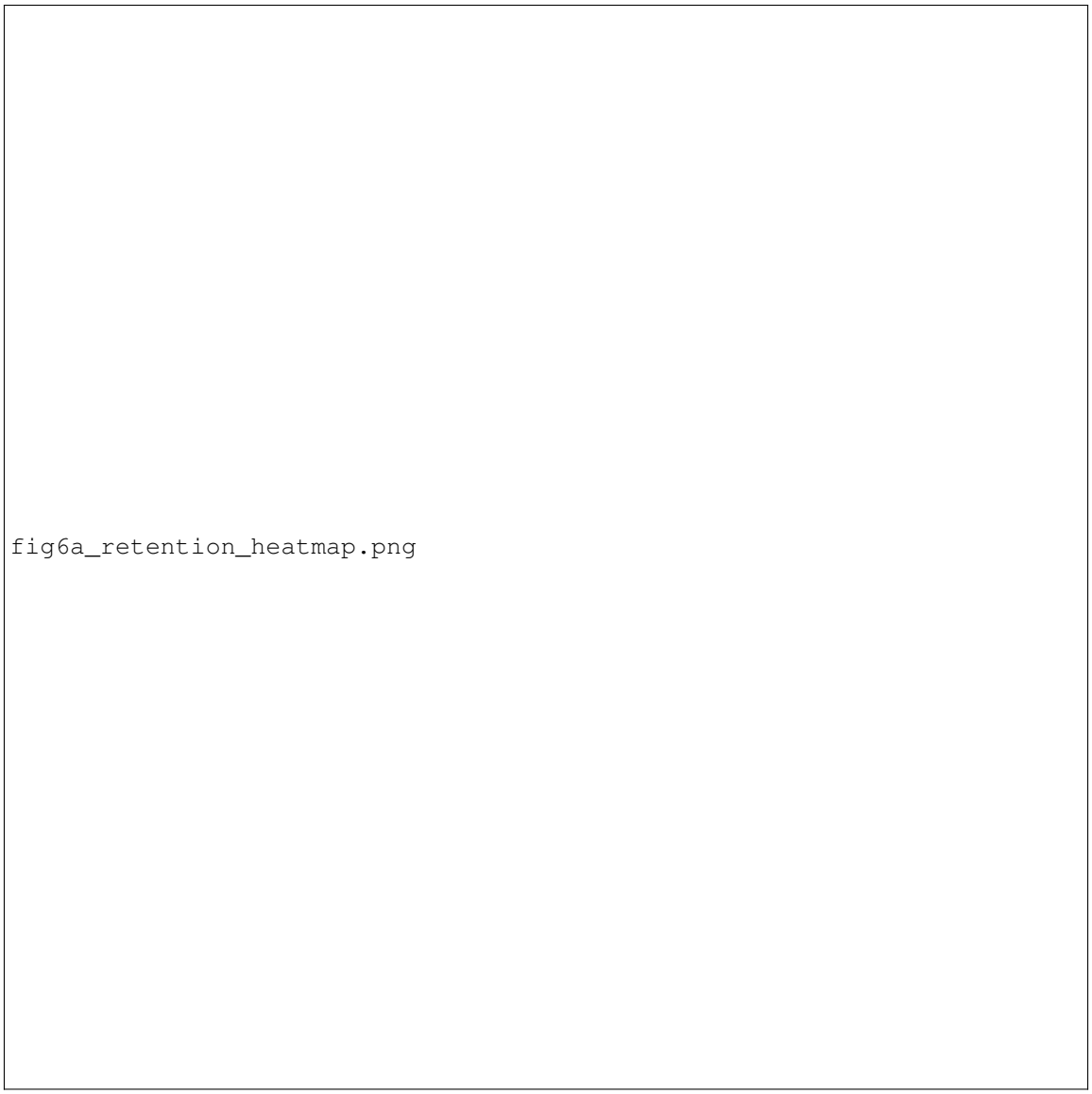


Figure 14: Event–time-bin retention rates $\hat{r}_{e,t}$ (archive coverage heatmap).

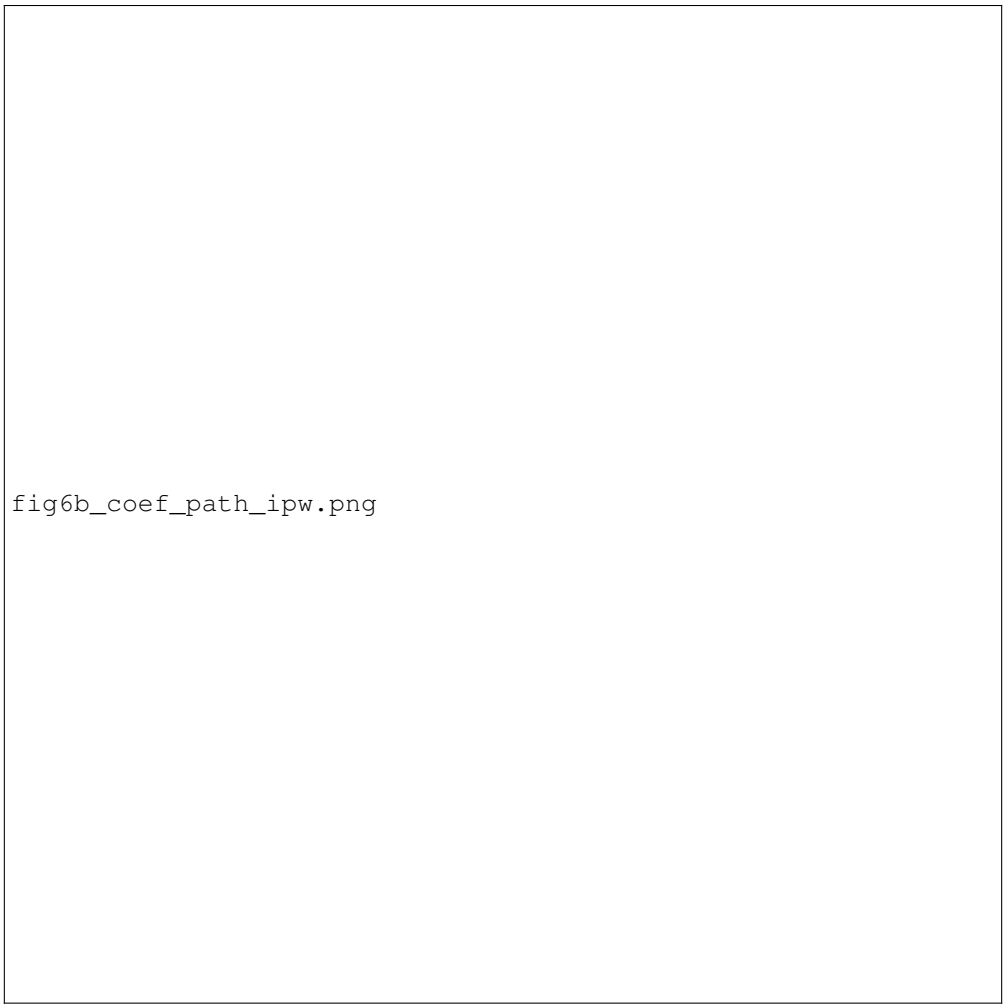


Figure 15: Coefficient sensitivity paths (unweighted vs IPW vs archive-only).

Integration: Combined models will relate content-level features, network position, and user roles to diffusion outcomes, enabling predictive and explanatory analyses of science communication dynamics. Specifically, we will model the diffusion probability as a function of content-level (e.g., sentiment, novelty), source-level (e.g., verification status, expertise), and structure-level (e.g., centrality in HIN) variables. The predicted outcomes include repost depth, breadth, and velocity. We will compare early fusion and late fusion strategies to assess how content and network features jointly predict diffusion outcomes.

Linking Online Events to Scientific Output: We define a high-engagement event as any Weibo post by an identified scientist that falls into the top decile of engagement (reposts, comments, or likes) within field–month. For each scientist i, we set the timestamp of the first qualifying event in 2015–2025 as t_0 (subsequent events are ignored for identification). We construct a balanced panel from $t_0 - 12$ to $t_0 + 24$ months, merging bibliometrics with social media histories. Outcomes include: (i) publication volume and field-normalized indicators; (ii) topic-distribution shifts from titles/abstracts (e.g., LDA/BERTopic); and (iii) collaboration metrics from co-authorship networks (new coauthors, degree/triad measures).

To improve comparability, we match exposed scientists to controls by discipline and career stage using pre-event trajectories (nearest-neighbor/propensity reweighting). Our main estimator is a two-way fixed-effects DiD:

$$y_{i,t} \;=\; \alpha_i \;+\; \lambda_t \;+\; \delta\big(Exposed_i \times Post_{i,t}\big) \;+\; X_{i,t}\beta \;+\; \varepsilon_{i,t},$$

with $X_{i,t}$ controls and clustered scientist-level standard errors. We implement stacked event-study estimators (Callaway–Sant'Anna) to handle staggered timing and perform placebo/permutation tests and Oster-style sensitivity bounds. We also exploit local randomization around platform thresholds (e.g., posts just above vs just below trending cutoffs) as quasi-experimental leverage where feasible.

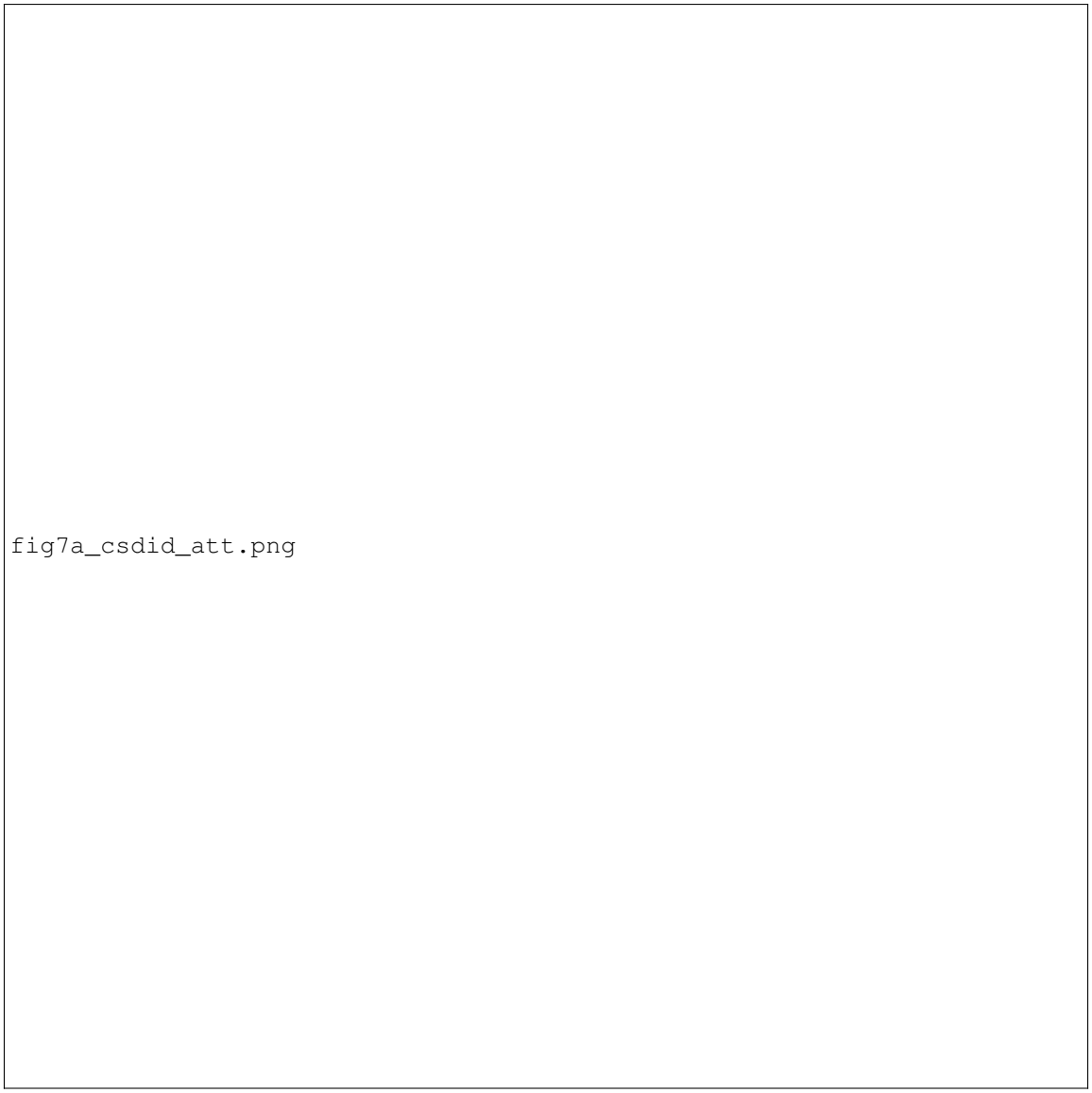


Figure 16: Callaway–Sant'Anna group-time ATT (topics/collab/productivity).

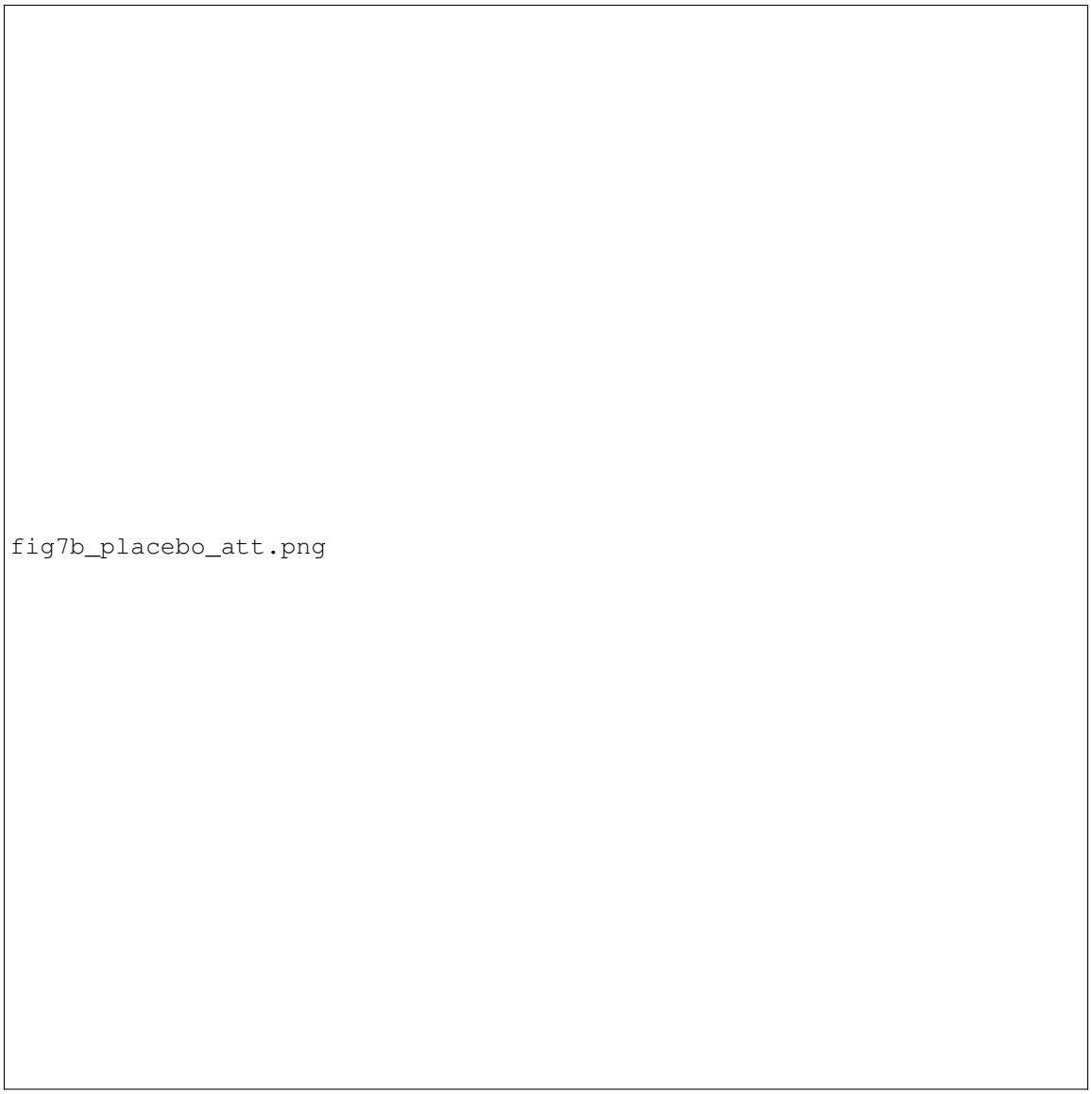


Figure 17: Placebo treatment timing: ATT centered around 0.

Mechanism Tests. To open the black box, we test mechanisms linking online experiences to scientific change. (1) Agenda salience: regress topic shifts on event-salience proxies (event peak intensity, media pickup counts) with scientist fixed effects. (2) Opportunity expansion: relate collaboration growth to new-follower influx and verified-user interactions during the event. (3) Institutional attention: track acknowledgments and topical alignment in subsequent publications as circumstantial evidence of downstream funding or institutional response. We instrument exposure intensity with exogenous timing near platform thresholds (e.g., trending cutoffs) and validate with permutation-based placebo events. Together, these tests provide a transparent window into how ephemeral online experiences may crystallize into enduring scientific outcomes. We proxy funding uptake via post-event increases in field-normalized topical alignment to new grant calls and via the appearance of funding-related acknowledgments; results are benchmarked against matched controls.

Threats to Identification and Robustness. Key risks include endogenous exposure, spillovers/interference, measurement error in exposure signals, and post-hoc deletion (survivorship) bias. We address these via:

1. cohort-specific stacked event studies with calendar-time fixed effects and discipline-by-time fixed effects;
2. exclusion windows for overlapping events within coauthor neighborhoods and neighbor-exposure controls;
3. common-support and pre-trend diagnostics with sensitivity bounds to unobserved confounding;
4. alternative exposure definitions (top 5%/15%; median split; moving-average windows);
5. a three-fold strategy for deletion bias: (i) cross-checking multiple archival sources and third-party datasets (Weiboscope, historical dumps); (ii) estimating snapshot-based retention models to reweight/adjust for survivorship (see above); (iii) reporting robustness restricted to archival/non-deleted subsets.

This project is designed for interoperability with existing Knowledge Lab pipelines. We will adopt unified diffusion indicators (depth, breadth, velocity) and release reproducible scripts, which can be directly applied to both Weibo and Western social media platforms. This design enables cross-cultural comparability and aligns with the Lab's commitment to open, reproducible computational frameworks.

Expected Challenges and Mitigation Strategies

Data Accessibility: API limitations and censorship may restrict data access. We will supplement with crawlers and seek data-sharing collaborations.
Semantic Ambiguity: Informal language and emojis may distort NLP results. We will use robust models, manual validation, and expert review for calibration.
Model Interpretability: To address GNN and causal model complexity, we will use explainability tools like SHAP and GNNExplainer.
Confounding Factors: Observational biases will be addressed through causal modeling frameworks, including covariate control and sensitivity analysis.
Ethics and Privacy: Only public data will be collected. Sensitive user information will be excluded or hashed; deleted-post recoveries will not be redistributed in raw form and aggregated outputs only will be shared publicly in accordance with IRB and platform terms. We have submitted an IRB application and will comply with platform Terms of Service; scraping will target publicly visible pages only and PII will be hashed or excluded.
Identity Matching: Linking Weibo accounts to actual scientists poses challenges, as some researchers may use pseudonyms or multiple accounts. We will validate identities through cross-referencing institutional profiles.
Data Completeness: Publication coverage across databases may be uneven, requiring triangulation of multiple sources to ensure robust data.
Attribution Limits: Shifts in research focus may arise from external factors (funding, societal events) rather than Weibo engagement. We will mitigate this by constructing control groups and acknowledging limitations.
Temporal Lag: Scientific outputs often appear 1–2 years after projects begin, complicating attribution to short-term online events. We will interpret findings cautiously and consider longer windows where data permit.

Anticipated Contributions and Novel Insights

This project is poised to generate novel insights into how science diffuses through algorithmically curated and state-regulated media environments. By integrating heterogeneous network modeling and causal inference, it moves beyond description to uncover the mechanisms driving science communication on Weibo.
Deliverables will include:

1. cleaned and anonymized data slices (subject to privacy constraints),
2. open-source code and reproducible pipelines (GitHub repository structured as `data/, src/, results/`),
3. staged outputs—pilot analysis, interim findings, and final publications.

Extended Data

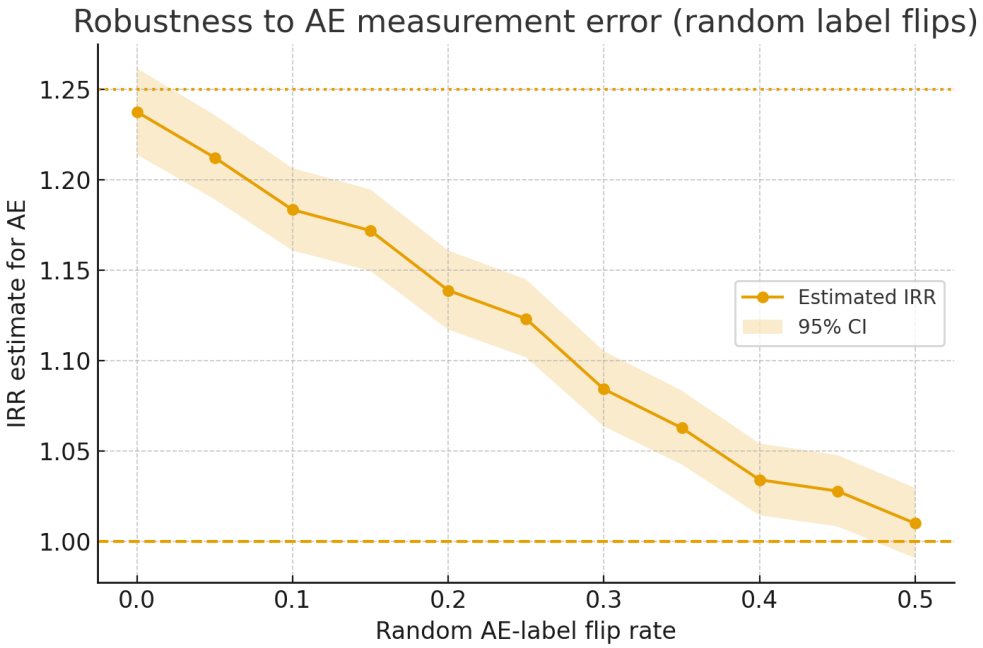


Figure 18: *(simulation)* Attenuation of the estimated AE effect under random misclassification. Points show the estimated incidence-rate ratio (IRR) for AE as the share of randomly flipped AE labels increases from 0 to 0.5; the band is the 95% CI across resamples. The dotted lines mark the no-effect benchmark (IRR=1) and a reference "true" effect (IRR≈1.25).

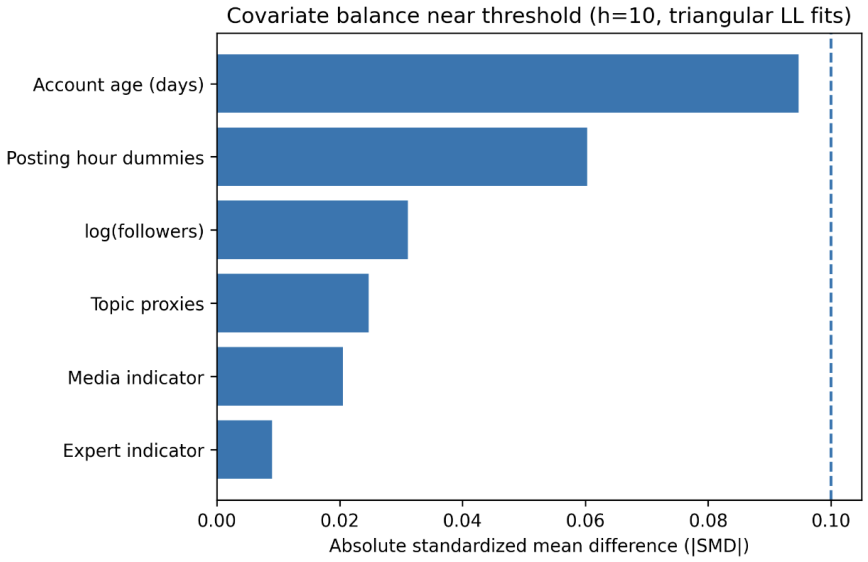


Figure 19: Covariate balance near the trending threshold. The figure reports absolute standardized mean differences (SMD) for pre-treatment covariates, estimated via local linear regressions with a triangular kernel within bandwidth $h = 10$ on each side of the cutoff. The horizontal dashed line marks $|\mathrm{SMD}| = 0.10$, a common benchmark for negligible imbalance (Austin 2009). Estimation follows robust RD practice (local-polynomial fits and data-driven bandwidth/binning) (Calonico, Cattaneo, and Titiunik 2014; Cattaneo, Idrobo, and Titiunik 2020).

ED1–ED3: Alternative definitions and thresholds for Cold_i and AE_i;
ED4: Count-model robustness (Poisson/NB2/ZIP/Hurdle; clustered SEs);
ED5: Multiple testing control (FDR);
ED6: Heterogeneity by discipline, followers, topic heat;
ED7: Network interference diagnostics;
ED8–ED9: Event-study windows and thresholds;
ED10: Ethics, de-identification, and data/code availability.

Table 4: Supplementary Table S1: Event pre-screen diagnostics (sources-verified time windows; counts are placeholders)

Event ID	Event name	Date range (verified)	Included	Initial sample	Archived sample	Estimated deletion rate (%)
E01	Jasic (worker-organizing)	2018-07–2018-09[a]	Y			
E02	996.ICU (tech labor)	2019-03–2019-06	Y			
E03	Foxconn controversies	2010-01–2011-12[b]	Y			
E04	Yue Yuen strike	**2014-04–2014-05**[c]	?			

[a] Core mobilization and detentions span July–September 2018; further detentions and clean-up actions extended into Nov/Dec 2018.
[b] Suicides and major attention peaked in early/mid 2010; broader labor-condition controversies persisted into 2011.
[c] The strike began mid-April 2014 (widely reported on 14 April) and continued into late April/early May.
Note: Public sources verify event existence and timing, but not the corpus counts. "Initial/Archived sample" and deletion rates ("–") are placeholders to be filled from our own crawls/archives (e.g., Weiboscope) using a reproducible diagnostics pipeline.

Provenance note. Event windows draw on reputable news and labor-research sources: Jasic detentions and censorship (China Labour Bulletin reports, 2018); the 996.ICU timeline (press coverage of March–April 2019); Foxconn suicides timeline (2010); and the Yue Yuen strike (April 2014). Corpus-level counts and deletion rates are not available from these sources and will be computed from our own Weibo crawls and third-party archives in the pilot.

References

- Brossard, D., & Scheufele, D. A. (2013). *Science, New Media, and the Public*. Science, 339(6115), 40–41. [link]
- Cui, H. & Kertész, J. (2021). *Attention dynamics on the Chinese social media Sina Weibo during the COVID-19 pandemic*. EPJ Data Science, 10(1), 8. [link]
- Evans, J. A. (2008). *Electronic publication and the narrowing of science and scholarship*. Science, 321(5887), 395–399. [link]
- McCombs, M. (2005). *A Look at Agenda-Setting: Past, Present and Future*. Public Opinion Quarterly, 69(4), 543–557. [link]
- Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Simon and Schuster. [link]
- Su, L. Y.-F., Akin, H., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2015). *Science News Consumption Patterns and Their Implications for Public Understanding of Science*. Journalism & Mass Communication Quarterly, 92(3), 597–616. [link]
- Vosoughi, S., Roy, D., & Aral, S. (2018). *The Spread of True and False News Online*. Science, 359(6380), 1146–1151. [link]
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). *A comprehensive survey on graph neural networks*. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4–24. [link]
- Youn, H., Strumsky, D., Bettencourt, L. M. A., Lobo, J., & Evans, J. A. (2014). *Invention as a combinatorial process: Evidence from US patents*. Journal of The Royal Society Interface, 13(123). [link]
- Yu, H., Xu, S., Xiao, T., Hemminger, B. M., & Yang, S. (2017). *Global science discussed in local altmetrics: Weibo and its comparison with Twitter*. Journal of Informetrics, 11(2), 466–482. [link]
- Zhao, Z., Zhao, J., Sano, Y., & Takayasu, M. (2018). *Fake News Propagate Differently from Real News Even at Early Stages of Spreading*. arXiv preprint. [link]
- Calonico, S., Cattaneo, M. D., Titiunik, R. (2014). *Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs*. *Econometrica*, 82(6), 2295–2326. [link]
- Cattaneo, M. D., Idrobo, N., Titiunik, R. (2020). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press (Elements in Quantitative and Computational Methods for the Social Sciences). [link]
- McCrary, J. (2008). *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test*. *Journal of Econometrics*, 142(2), 698–714. [link]
- Austin, P. C. (2009). *Balance Diagnostics for Comparing the Distribution of Baseline Covariates Between Treatment Groups in Propensity-Score Matched Samples*. *Statistics in Medicine*, 28(25), 3083–3107. [link]