entropy

MDPI

Article

Local Weighted-Degree Entropy Serves as a Diagnostic and Routing Signal for Temporal GNN Aggregation

Wei He

Xi'an Jiaotong University International Research Center for Information Philosophy (XJTU-ICPI), School of Humanities and Social Sciences, Xi'an Jiaotong University, Xi'an 710049, China; hw985211@stu.xjtu.edu.cn

Abstract

Temporal link prediction is usually scored by a single outer-test average, which treats predictive risk as exchangeable across dyads. Contact and online-learning streams violate that assumption: recurrent motifs coexist with sparse, weakly recurrent edges. We study whether the relative out-of-sample risk of snapshot message passing is heterogeneous across local weighted-degree-entropy states, and whether an entropy-conditioned router yields selective rather than uniform risk reductions. On SocioPatterns school/workplace contacts, BenchTemp massive open online course (MOOC) and contact-like panels, and synthetic temporal graphs, we compare a calendar–recency logistic baseline, protocol-matched temporal graph neural network (GNN) controls, an ungated temporal graph convolutional network (GCN), and an entropy-gated dual-head model under a chronological 50/25/25 split. Normalized local weighted-degree entropy serves as both an evaluation stratum and a soft gate between message-passing and calendar–recency logits. In a fixed medium protocol (12 units), gating improves mean outer-test log-loss relative to the ungated GCN on school ($\Delta = -0.015$) and workplace ($\Delta = -0.003$), with school gains concentrated in the highest-entropy quintile (mean $\Delta = -0.071$). These gains are relative to a weaker ungated GCN; a calendar–recency logistic baseline often remains competitive or better in absolute log-loss. Heterogeneous-exponential synthetics favor gating, whereas memory-free and homogeneous-exponential controls can produce harm; multi-seed and transfer screens bound stability. The contribution is diagnostic: entropy-conditioned errors matter when risk is uneven, but the effect is neither universal across domains nor stable under all generators.

Keywords: local weighted-degree entropy; temporal graph neural network; entropy gating; interaction-event forecasting; SocioPatterns; BenchTemp; conditional predictive risk; temporal networks

Check for updates

Academic Editor: Tiziano Squartini

Received: 19 September 2026
Revised: 15 November 2026
Accepted: 16 November 2026
Published: 18 November 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Predicting the next interaction on a temporal network supports epidemiology, workplace design, and online learning analytics [1–4]. Contact sensors and platform logs yield streams of dyadic events whose short-term recurrence is highly structured, yet most published evaluations still collapse model quality into one network-wide score. That habit rewards predictors that average well everywhere, even when compressibility concentrates in a few recurrent motifs and the exploratory periphery remains hard to forecast [5,6].

Despite that heterogeneity, temporal link predictors are usually ranked by a single outer-test average [7–9]. A network-wide score is convenient for benchmarks, but it does not ask where conditional predictive risk concentrates, nor whether neighborhood aggregation

helps the same dyads that carry the bulk of the loss. When recurrent and exploratory structure coexist, that omission is costly: a large gain on a minority of high-entropy edges can vanish in a global mean, and a global gain can be driven by a narrow stratum that does not generalize.

Graph neural networks (GNNs) address relational dependence by aggregating messages over local neighborhoods [10–12]. Temporal extensions maintain node memories, evolve snapshot embeddings, or induce continuous-time representations [13–17]. Parallel work inserts Shannon or structural entropy into GNN pipelines as feature weights, attention coefficients, information-bottleneck objectives, or mixture-of-experts regularizers [18–29]. Those designs treat entropy mainly as an internal knob for raising a global task score.

What remains under-examined for temporal link prediction is whether outer-test loss itself varies systematically with a local information-theoretic state of the subgraph surrounding each candidate dyad, and whether an entropy-conditioned gate helps only where neighborhood aggregation is informative—without forcing a uniform global gain or claiming a new state-of-the-art backbone. Answering that question requires (i) a reproducible local entropy on rolling history, (ii) separable message-passing (MP) and calendar–recency pathways, and (iii) reporting protocols that stratify outer-test error rather than only averaging it.

This paper addresses a single research question. On temporal interaction networks, is the relative out-of-sample risk of snapshot message passing heterogeneous across local weighted-degree-entropy states; if so, can an entropy-conditioned router produce selective rather than uniform risk reductions relative to uniform aggregation, and how do those patterns change with network type and generative memory? The distribution of conditional predictive risk, the gating mechanism, and boundary conditions under different generators are analysis dimensions of that question, not three independent projects.

The contributions are threefold. (i) We define a normalized local weighted-degree Shannon entropy on the rolling ego-union subgraph for each candidate dyad and use that quantity as an outer-test stratification variable, with explicit caveats that $H_e = 1$ marks degree balance rather than exploratory uncertainty, and that empty history is coded as $H_e = 0$ by convention. (ii) We introduce an entropy-gated temporal graph convolutional network (GCN) that mixes a message-passing logit with a calendar–recency logit through a soft gate driven by the same local entropy; we emphasize that the gated model differs from the ungated GCN by both the gate and the additional calendar head, so gated-versus-ungated Δ is not a pure entropy-only contrast. (iii) We evaluate global and stratified log-loss on public school and workplace contact streams, BenchTemp interaction panels, and synthetic controls with known memory structure, and we place the gated model against protocol-matched classical and temporal-GNN baselines to isolate when gating helps and when memory-free regimes form a boundary. Throughout, we distinguish our diagnostic claim from architecture-only entropy–GNN papers that optimize a single global metric [26–28]. On SocioPatterns contacts, calendar–recency logistic often has lower absolute log-loss than either GCN variant; the dual-head GCN is used to stratify risk and to test entropy-conditioned routing (Section 5.1). Multi-seed contact intervals for Δ include zero (Section 4.18), so setting-level means are reported as descriptive summaries.

The remainder of the paper is organized as follows. Section 2 reviews temporal link prediction and entropy-aware graph learning. Section 3 details data, entropy, models, and chronology. Section 4 reports experiments. Sections 5 and 6 discuss implications and conclude.

2. Related Work

2.1. Temporal Link Prediction on Interaction Networks

Classical link prediction combines topological similarity with time-decayed edge features [7–9]. On continuous-time event streams, Hawkes and related point-process models capture self-excitation and mutual excitation among dyads [2,21]. Those models are attractive when the scientific interest is in intensity dynamics, but they do not by themselves provide a graph-neighborhood representation for sparse candidate sets. Surveys of dynamic and temporal graph representation learning summarize snapshot GNN stacks, continuous-time memories, and inductive architectures [1,17,30]. Representative systems include JODIE-style trajectory models, temporal graph networks with memory modules, EvolveGCN-style recurrent weight evolution, and inductive continuous-time embeddings [13–16].

Public contact corpora from the SocioPatterns collaboration remain standard face-to-face benchmarks [3–6,31]. Those deployments (high-school, workplace, and related proximity settings) are widely used both as epidemiology contact proxies and as temporal-network benchmarks; published pipelines typically bin contacts, construct chronological folds, and report network-wide area under the ROC curve (AUC)/area under the precision–recall curve (AUPRC) or proper scores without entropy-conditioned strata. Educational interaction benchmarks such as BenchTemp supply heterogeneous learner–item streams that stress bipartite structure [32]. BenchTemp further packages contact-like and online interaction panels (e.g., Contacts, SocialEvo, Wikipedia, Reddit, LastFM) under a common evaluation interface, which is why we reuse its panel construction for volume and transfer screens rather than treating massive open online course (MOOC) panels as the only non-contact control. On the model side, continuous-time and snapshot temporal-GNN benchmarks commonly feature JODIE, temporal graph attention (TGAT)/temporal graph network (TGN)-style memories, EvolveGCN, and related inductive encoders [13–16]; our protocol-matched lite controls borrow those design motifs under fixed width/depth, but we do not claim a faithful public-leaderboard reproduction. Across these settings, published evaluations typically report a ranking score (AUC, AUPRC) or a proper scoring rule on a held-out time interval. They rarely condition outer-test error on a local information-theoretic state of the subgraph surrounding each dyad. Community-detection and epidemic analyses on the same contact data emphasize temporal motifs and burstiness [33,34], which again suggests that predictability need not be spatially uniform even when a single average loss looks stable.

2.2. Entropy Measures and Information-Constrained Graph Learning

Shannon's discrete entropy and its network extensions provide operational summaries of uncertainty and concentration [35–38]. Structural entropy, in particular, has been used to characterize hierarchical organization and to guide multi-network accessibility analyses [18,19]. Transfer entropy quantifies directed predictive information between processes and has been combined with Hawkes intensities to measure contagion on directed weighted networks [20,22]. Information-bottleneck formulations constrain representations so that only task-relevant bits are retained [23]; graph information bottleneck and related clustering objectives transplant that idea to relational data [24,25].

These tools appear throughout recent *Entropy* network papers, usually as descriptive indices of a static or multilayer graph, as learning objectives, or as diagnostics of cascade risk. What they less often provide is an outer-test stratum for short-horizon link prediction: a dyad-level, history-strict quantity that can be recomputed for every prediction time and then used both to gate a model and to slice the test loss.

2.3. How Prior Entropy–GNN Work Differs from This Paper

Several recent *Entropy* articles already couple entropy with GNNs in ways that are architecturally close to ours yet scientifically distinct. Xi and Cui weight degree-based node features with information entropy before message passing for influential-node identification [26]. Wang uses structure-entropy-based attention to aggregate sensor-graph neighbors for industrial multivariate anomaly detection [27]. Kang et al. propose a spatio–spectral adaptive mixture-of-experts GNN with global structural priors for node classification [28]. Chen proposes adaptive attention trained with a cross-entropy loss [29]. Related lines include GTAD-style graph–temporal anomaly detectors [39,40] and conditional information-bottleneck graph clustering on dynamic networked systems [25]. Surveys of over-smoothing [12], together with entropy-preserving augmentation against over-smoothing [41], remind us that naive deep aggregation can destroy local distinguishability, which is one motivation for not treating message passing as uniformly helpful.

Shared premise. Entropy can usefully enter graph learning.

Shared limitation for our research question. Those works optimize or report a global task metric. Entropy reweights aggregation, features, or expert selection; it is not used as a primary lens for *where* predictive error lives on a temporal interaction network.

This paper. We (i) compute local weighted-degree entropy for candidate dyads in a rolling window, (ii) stratify outer-test log-loss by entropy quintiles and by marginal message-passing gain, and (iii) use entropy only as a soft gate between a message-passing head and a calendar–recency head under chronological link prediction on contact, MOOC, and synthetic streams. Success is defined by interpretable stratified deltas and boundary regimes, not by claiming a new state-of-the-art GNN backbone.

3. Materialsand Methods

3.1. Problem Setting and Estimand

Let $\mathcal{E} = \{(u_k, v_k, t_k)\}$ be a stream of undirected interaction events among a finite node set V. We discretize time into bins of width Δt and write $y_{e,t} \in \{0, 1\}$ for whether candidate dyad $e = (u, v)$ is active in bin t. Given history $\mathcal{H}_{<t}$ and calendar covariates c_t, the model outputs a probability $\hat{p}_{e,t} = \sigma(\ell_{e,t})$ for each candidate in a fixed edge set E. The estimand is predictive: we ask how well $\hat{p}_{e,t}$ scores the outer-test events under a proper scoring rule. Because candidates are restricted to frequent training dyads, the task is short-horizon interaction-event forecasting among previously observed pairs (edge reactivation), not unrestricted cold-start link discovery. We do not claim causal effects of gating on social mechanisms.

Primary metric: mean binary log-loss on the outer-test block,

$$\mathrm{LL} = \frac{1}{|\mathcal{T}_{\text{test}}||E|} \sum_{t \in \mathcal{T}_{\text{test}}} \sum_{e \in E} [-y_{e,t} \log \hat{p}_{e,t} - (1 - y_{e,t}) \log(1 - \hat{p}_{e,t})]. \tag{1}$$

Secondary metric: AUPRC, which is more informative than AUC under extreme class imbalance. When comparing gated and ungated GCNs, we report

$$\Delta = \mathrm{LL}_{\text{gated}} - \mathrm{LL}_{\text{ungated}}, \tag{2}$$

so that negative Δ favors gating.

3.2. Data Sources and Preprocessing

3.2.1. SocioPatterns Contact Networks

We use the 2012 high-school and 2015 workplace deployments [4,6,31]. Raw proximity records are collapsed to episode onsets so that duration does not dominate the temporal signal, then binned at $\Delta t = 5$ min into binary edge events. Rolling folds define outer-test days; each

inferential unit retains the 400 most frequent training dyads as the candidate set E. Inferential units are outer-test days. These scenes mix scheduled, repeatedly activated contacts with sparse encounters and therefore supply a natural testbed for entropy-conditioned predictability.

3.2.2. BenchTemp Interaction Streams

Learner–item interactions from BenchTemp [32] are binned at $\Delta t = 30$ min into panel-windows for the MOOC control. Each panel keeps about one thousand frequent learner–item pairs. Local entropy is computed on the bipartite ego union (learner neighborhood union item neighborhood). As contact-like volume checks, we additionally reuse the same chronological panel construction on BenchTemp Contacts and SocialEvo streams. Primary domains are predefined as undirected proximity/face-to-face contact networks (school, workplace, and contact-like BenchTemp streams). Heterogeneous online bipartite panels (e.g., Wikipedia/LastFM-style logs) are retained as cross-domain transfer checks: under the same protocol their gated-versus-ungated signs attenuate or reverse, which we treat as a challenge to transfer rather than as a reason to drop them from the scientific record (Supplementary Materials summaries report those screens).

3.2.3. Synthetic Temporal Graphs

On a fixed Watts–Strogatz skeleton [42], we generate event streams of length $T = 240$ bins under four mechanisms: memory-free Bernoulli activation, homogeneous exponential waiting, Hawkes self-excitation [21], and heterogeneous exponential waiting. The synthetic layer holds topology fixed while varying generative memory, which is essential for the boundary dimension of the research question.

3.2.4. Chronological Split

All streams use the same split: first 50% of bins for fitting, next 25% for validation (used to freeze gate hyperparameters and marginal-gain quintile edges; not used for validation-based early stopping under the locked medium protocol), and final 25% as the unique outer test. Gate thresholds and training hyperparameters are never chosen on the outer-test block. The fixed medium protocol uses 12 units (school × 2, workplace × 3, MOOC × 3, synthetic × 4). A fuller sweep with more school/workplace outer-test days and all available MOOC panels ($n = 27$ units including synthetics), together with the extended BenchTemp panels above, is reported as robustness evidence and does not redefine the locked medium tables in Section 4.

3.3. Local Subgraph Entropy

For each time t and candidate dyad $e = (u, v)$, form the weighted multigraph of events whose timestamps fall in the half-open window $[t - W, t)$ with $W = 12$ bins. Let $N(u)$ and $N(v)$ be the sets of nodes that co-appear with u and v inside that window, and define the ego-union node set

$$S_e = \{u, v\} \cup N(u) \cup N(v). \tag{3}$$

Write d_i for the weighted degree of node $i \in S_e$ in the window graph and $p_i = d_i / \sum_{j \in S_e} d_j$. The normalized Shannon entropy is

$$H_e(t) = \frac{-\sum_{i \in S_e} p_i \log p_i}{\log |S_e|}, \tag{4}$$

with the conventions $H_e(t) = 0$ if $|S_e| \leq 1$ or if the total mass is zero. Normalization by $\log |S_e|$ maps the index to $[0, 1]$ and makes dyads with different ego sizes comparable. Crucially, the window ends strictly before t, so entropy used at prediction time does not leak outer-test labels. Two boundary cases matter for interpretation. A repeatedly activated

isolated dyad with balanced endpoint mass yields $H_e \to 1$, so "high entropy" here means local weighted-degree balance, not necessarily exploratory unpredictability. Empty recent history is also coded $H_e = 0$ and therefore receives a relatively large gate weight on message passing ($g(0) \approx 0.858$), which is a modeling convention rather than a claim that empty neighborhoods are concentrated. We therefore call H_e a local weighted-degree/degree-balance entropy throughout.

We intentionally choose a degree-based Shannon form rather than a spectral or community-partition structural entropy [18]. The degree form is computationally inexpensive enough to recompute for every (e, t), transparent to readers outside network science, and sufficient to separate concentrated star-like neighborhoods from more balanced local mass. Stronger structural entropies remain compatible with the same stratified evaluation protocol. Physically, $H_e(t)$ is an instantaneous degree-balance thermometer on the recent ego union: values near 0 correspond to mass concentrated on one or a few partners (star-like or empty-coded history), whereas values near 1 correspond to more evenly shared recent interaction mass among neighbors. In contact systems, this often tracks schedule-driven mixing versus repeated dyadic meetings; it is *not* a direct measure of semantic novelty or exploratory intent. The gate therefore uses H_e as an operational proxy for "how diffuse is recent neighborhood mass," not as a claim that high entropy equals unpredictable social exploration.

3.4. Models

Figure 1 sketches the architecture. All neural models share the same candidate set, calendar features, and chronological split, which holds the evaluation protocol fixed. The main gated-versus-ungated contrast still differs by both the entropy mixer and the additional calendar–recency head; Section 4.8 therefore isolates routing under shared dual heads, while the primary Δ should be read as a dual-head routing contrast rather than a pure information-theoretic ablation.

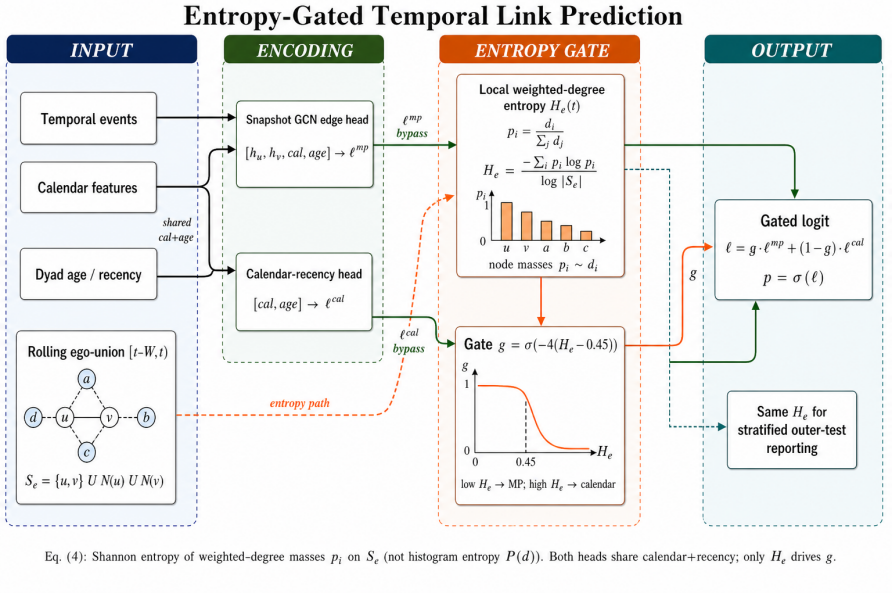


Figure 1. Architecture of the entropy-gated temporal link predictor. Arrows follow the forward computation: rolling ego-union history yields the normalized local weighted-degree Shannon entropy $H_e(t)$ of Equation (4) (node masses $p_i \propto d_i$ on S_e; not a degree-histogram entropy over $P(d)$), which soft-gates the mix of snapshot graph convolutional network (GCN) message-passing (MP) logits ℓ^{mp} and calendar–recency logits ℓ^{cal}. The history window ends strictly before the prediction bin.

3.4.1. Logistic Baseline

A logistic regression uses calendar Fourier features of bin index together with an exponential recency feature of dyad age,

$$a_{e,t} = \begin{cases} \exp(-\text{age}_{e,t}/36), & \text{age}_{e,t} < \infty, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where age is measured in bins since the last activation of e. This baseline has no neighborhood aggregation and therefore anchors how far calendar–recency alone can go.

3.4.2. Ungated Temporal GCN

At each prediction time t, we build a snapshot adjacency strictly from bins $\tau < t$ (implementation: the cumulative event history up to but not including bin t), never from the target bin t whose labels $y_{e,t}$ are being scored. We run a two-layer GCN [10] with hidden width 32. Node inputs concatenate a learned embedding with simple degree-style features computed on the same past-only snapshot. An edge head maps the concatenation of endpoint embeddings, calendar features, and recency to a message-passing logit $\ell^{\text{mp}}_{e,t}$.

3.4.3. Entropy-Gated Temporal GCN

A second head maps calendar and recency features alone to a calendar logit $\ell^{\text{cal}}_{e,t}$. The gated score mixes the two pathways,

$$\ell_{e,t} = g_{e,t}\,\ell^{\text{mp}}_{e,t} + (1 - g_{e,t})\,\ell^{\text{cal}}_{e,t}, \qquad g_{e,t} = \sigma\big(-4\big(H_e(t) - 0.45\big)\big). \tag{6}$$

Because $\sigma(-4(H - 0.45))$ is decreasing in H, low local entropy favors message passing and high local entropy favors the calendar–recency path. Numerically, $g(0) \approx 0.858$ and $g(1) \approx 0.100$; we avoid the looser "$g \gtrsim 0.88$" wording. The midpoint 0.45 and slope 4 were fixed on validation behavior in preliminary runs and then held constant across all reported units; they are not re-tuned on outer test. We keep the gate parametric rather than end-to-end learned so that the same scalar H_e remains an inspectable diagnostic stratum; Section 4.11 reports a midpoint/slope grid, and Section 4.20 evaluates an end-to-end learned (c, s) parameterization as a sensitivity check, and Section 4.8 compares entropy mixing to constant, shuffled, and activity mixers under shared dual heads (the same trained heads are fixed across mixers; only the mixer changes). Empty-history coding $H_e = 0$ therefore routes mostly toward message passing by convention ($g(0) \approx 0.858$); we treat that as an explicit modeling choice, not as evidence that cold dyads should prefer aggregation. The design makes the gate an explicit, inspectable function of the same quantity used for stratification. Because stratification and gating share H_e, gated-minus-ungated differences are partly mechanical; we therefore also report validation marginal-gain strata and treat absolute logistic performance as a separate baseline question.

3.4.4. Marginal Message-Passing Gain

To separate "low entropy" from "message passing is useful," we also compute, on the validation block, a per-dyad marginal gain: the mean reduction in binary cross-entropy when scoring with ℓ^{mp} instead of ℓ^{cal}. Quintiles of that validation gain become a second stratification axis on the outer test. If entropy-gated improvements tracked only low-entropy bins but not high marginal-gain bins (or conversely), the two proxies would be telling different stories.

3.4.5. Matched-Capacity Comparison Baselines

Related work mentions stronger temporal backbones, but the scientific target here is to isolate entropy-conditioned selection rather than to claim a new state-of-the-art encoder.

We therefore add a matched-capacity comparison suite under the same candidates, features, and chronological split:

- Time-aware Common Neighbors and Preferential Attachment scores, calibrated with logistic regression on calendar and recency features [7];
- A multilayer perceptron (MLP) edge head that uses node embeddings with calendar/recency but no graph convolution;
- A lightweight node-memory model (TGN-style exponential memory update, not a full Temporal Graph Network [14]);
- A snapshot graph attention network (GAT) [11] with the same hidden width as the ungated GCN;
- Lightweight EvolveGCN-O and T-GCN controls that evolve a compact weight code or a per-node gated recurrent unit (GRU) state across snapshots [15].

These baselines are intentionally protocol-matched and approximately capacity-matched. They answer whether simply swapping aggregation style or adding a recurrent temporal state removes the need for entropy-conditioned reporting under our shallow-snapshot protocol; they are ablation-style controls, not faithful reproductions of full TGN/EvolveGCN/T-GCN systems, and they do not replace a public-implementation temporal-GNN benchmark.

3.5. Training Protocol and Implementation Notes

Models are trained with binary cross-entropy (BCE) on the fit block. For medium-protocol contact and MOOC units, we use a small number of epochs (typically 4–5) and a temporal train stride of 3–4 bins to keep wall-clock time manageable while still covering the day; synthetic units use comparable budgets. Optimization uses the Adam optimizer with standard library defaults. Implementation details used for all locked runs: PyTorch 2.5 (Meta Platforms, Inc., Menlo Park, CA, USA) default parameter initialization (Embedding/Linear defaults; no custom Xavier override); Adam with learning rate 10^{-2} and library β defaults; fixed epoch budget without validation-based early stopping in the medium protocol (the validation block freezes gate hyperparameters and marginal-gain quintile edges, but does not truncate training); no class re-weighting or negative sampling beyond the fixed frequent-dyad candidate set (extreme imbalance is handled by reporting log-loss and AUPRC jointly); candidate construction keeps the top-400 training-frequency contact dyads per fold (about 1000 frequent learner–item pairs per MOOC panel). All neural models share this candidate set and chronological split. Snapshot GCNs are shallow by design: the scientific claim concerns entropy-conditioned selection between pathways, not depth. Ungated and gated models share width and depth for the message-passing encoder, but the gated model additionally includes a calendar–recency head and the entropy mixer; "matched capacity" therefore does not mean an identical parameter set. The main gated-versus-ungated Δ is therefore a dual-head routing contrast. Section 4.8 isolates the mixer by freezing jointly trained heads and swapping only post hoc gates (constant/shuffled/entropy/activity); we interpret those shared-head controls as the cleaner entropy-routing check under a fixed dual-head parameter set (edge-head and calendar-head widths remain as in Table 1, so the heads are not width-matched to each other), not as unfinished work.

Table 1 lists the fixed medium-protocol modeling choices. They are intentionally conservative. Increasing depth or width would change absolute losses and might improve the ungated GCN on some days, but would also confound the isolation of the entropy gate. Our reporting standard is therefore protocol-matched encoders plus stratified deltas, not an unrestricted architecture search.

Table 1. Fixed medium-protocol modeling choices (shared by ungated and gated graph convolutional networks (GCNs) unless noted). MLP: multilayer perceptron; AUPRC: area under the precision–recall curve; MOOC: massive open online course.

Choice	Value
Snapshot GCN depth/hidden width	2/32
Edge-head MLP	$2h + c + 1 \to 64 \to 1$
Calendar-head MLP (gated only)	$c + 1 \to 32 \to 1$
Initialization	PyTorch 2.5 Embedding/Linear defaults
Optimizer	Adam, lr $= 10^{-2}$, library β defaults
Epochs/train stride (medium)	4–5/3–4 bins (fixed budget)
Early stopping	Not used for medium locked runs
Class imbalance	No reweighting; log-loss + AUPRC
Entropy window W	12 bins
Gate midpoint/slope	0.45/4
Candidate dyads (contact)	top-400 by training frequency
Candidate pairs (MOOC panel)	≈1000 frequent pairs
Chronological split	50%/25%/25%
Contact bin width	5 min
MOOC bin width	30 min
Primary/secondary metrics	log-loss/AUPRC

Computation at Test Time

For every outer-test bin t and every candidate $e \in E$, we (i) rebuild the rolling window graph on $[t-W, t)$, (ii) compute $H_e(t)$ from Equation (4), (iii) encode the past-only snapshot $A_t^- = \sum_{\tau<t} A_\tau$ (equivalently, history before t) with the trained GCN, and (iv) form either ℓ^{mp} (ungated) or the mixture in Equation (6) (gated). Stratified metrics reuse the stored per-prediction tuples $(y_{e,t}, \hat{p}_{e,t}, H_e(t))$ so that quintile boundaries can be recomputed from the same artifact. Marginal-gain strata are frozen from the validation block and then applied to outer-test predictions, preventing the task proxy from peeking at test labels.

Negative evidence is admitted by construction. On memory-free synthetics, neighborhood structure in the recent window can be largely irrelevant noise, so forcing or softly preferring message passing may hurt under some seeds. We treat that possibility as a scientifically useful boundary rather than as a failure to optimize a single global benchmark.

3.6. Reproducibility and Inferential Units

Each SocioPatterns outer-test day is an inferential unit: preprocessing, candidate selection, and chronological split are repeated per day so that day-level dependence does not masquerade as an inflated sample of dyad–bin pairs. Each MOOC panel-window is likewise an inferential unit. Synthetic generators contribute one aggregate stream per mechanism in the fixed medium protocol, which is enough to sketch sign patterns but not enough for within-mechanism confidence intervals; a five-seed check on selected generators is reported later as robustness evidence. Medium-protocol tables in this manuscript are held fixed; Supplementary Materials sweeps with additional outer-test days and MOOC panels do not redefine the reported numbers here.

3.7. Experimental Design

We report nine analyses that map onto the distribution, mechanism, and boundary dimensions:

1. Global outer-test comparison among logistic, ungated GCN, and entropy-gated GCN.
2. Conditioning of Δ on local-subgraph-entropy quintiles.
3. Conditioning of Δ on validation marginal-gain quintiles.

4. Scene split between school and workplace contacts.
5. Heterogeneous MOOC control.
6. Synthetic generator sensitivity (memory-free, uni-exp, Hawkes, het-exp).
7. Ablations that remove the gate, message passing, or the age feature (school-day mean plus multi-fold contact units).
8. Matched-capacity comparison against classical topological scores and temporal-GNN lite controls.
9. Extended BenchTemp panels and a fuller multi-day/panel robustness sweep.

Primary reported numbers for the gated-versus-ungated claim in Section 4 use the medium protocol described above. Comparison-suite and extended-panel numbers use the same chronological protocol and are reported as complementary evidence. Replication materials are summarized in the Data Availability statement.

4. Results

Unless noted, numbers are setting-wise means from the medium protocol (12 units). Negative Δ means gated log-loss is lower (better) than ungated GCN. Day-level signs are not unanimous (school $+0.0015/-0.0319$; workplace mixes negative and positive days), so setting means should be read as fragile averages over few units rather than as stable within-scene laws. Comparison-suite, fuller-sweep, and extended-panel tables use the same chronological split but are separate reruns; absolute losses need not match the locked medium table cell-by-cell.

4.1. Global Outer-Test Comparison

Table 2 and Figures 2 and 3 summarize mean outer-test log-loss and the gated-versus-ungated Δ. Contact scenes show mean gated improvements: school $\Delta = -0.0152$ and workplace $\Delta = -0.0028$. MOOC remains small in absolute terms (mean $\Delta = -0.0018$), so we read MOOC as an attenuated control rather than as a second contact-like success case. The logistic baseline is competitive on several contact folds and often stronger than the ungated GCN in absolute log-loss, which is expected when calendar structure and dyad recency already explain much of the short-horizon signal; the scientific contrast of interest is therefore gated versus ungated GCN under a shared shallow encoder (with the gated model adding a calendar head), not a claim that shallow GCNs dominate logistic baselines on every day.

Reading the three models together clarifies why that contrast matters. On school and workplace, mean ungated GCN log-loss exceeds the logistic baseline (school: 0.0538 versus 0.0354; workplace: 0.0775 versus 0.0625), even though both models see the same calendar and recency features. The gap indicates that a shallow snapshot GCN can *inflate* proper scores when message passing is applied uniformly: neighborhood aggregation adds variance on dyads whose recent ego unions are weakly informative at a five-minute horizon. Entropy gating does not fully correct that degradation globally—gated school loss (0.0385) remains above logistic—but it closes most of the GCN–logistic gap by re-weighting toward the calendar–recency head where local subgraphs are diffuse. On synthetics, the ranking permutes: memory-free data reverse the contact pattern (logistic 0.2069, ungated 0.0834, gated 0.1232), because a memory-free generator makes neighborhood mass actively misleading while calendar structure is unhelpful; het-exp is the opposite extreme (ungated 0.2649, gated 0.1786), where heterogeneous waiting times create pockets where mixing pathways helps. These permutations are themselves evidence that "best model" is not a single label across generators.

Table 2. Mean outer-test log-loss by setting (medium protocol). Negative Δ favors the entropy-gated model over the ungated GCN.

Setting	n	Logistic	Ungated GCN	Gated	Δ
School	2	0.0354	0.0538	0.0385	−0.0152
Workplace	3	0.0625	0.0775	0.0747	−0.0028
MOOC	3	0.0112	0.0148	0.0129	−0.0018
Memory-free	1	0.2069	0.0834	0.1232	+0.0398
Uni-exp	1	0.1782	0.1900	0.2508	+0.0608
Hawkes	1	0.1996	0.3369	0.2384	−0.0985
Het-exp	1	0.1669	0.2649	0.1786	−0.0863

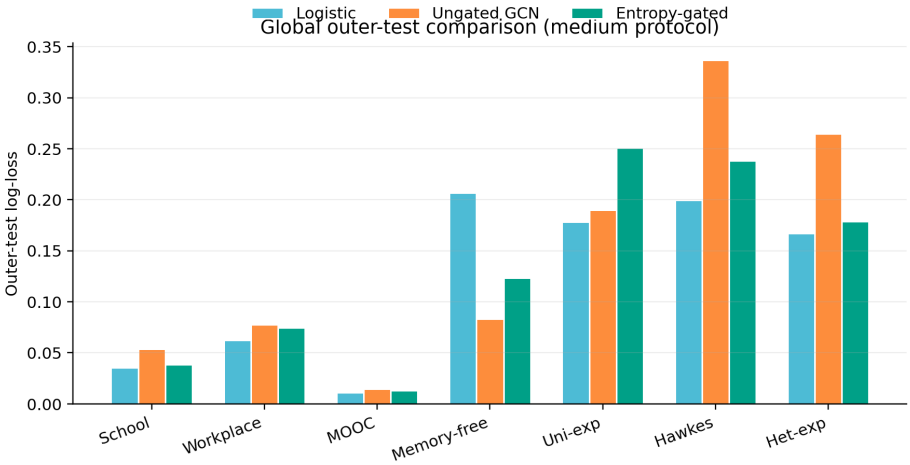


Figure 2. Global outer-test log-loss: logistic baseline, ungated GCN, and entropy-gated mix across settings in the medium protocol.

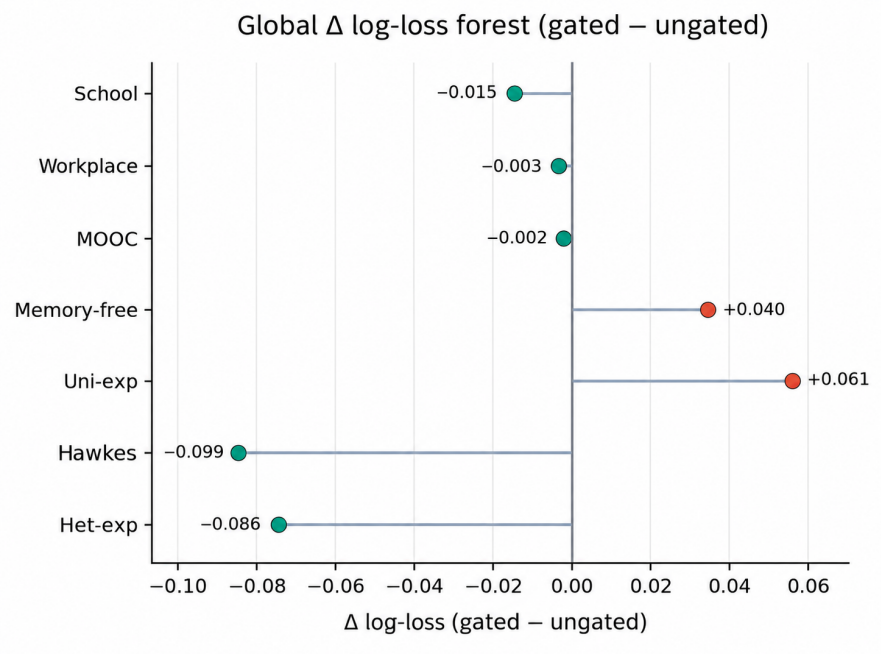


Figure 3. Global Δ log-loss (gated minus ungated) as a forest/lollipop plot by setting. Negative values (teal) favor entropy gating; positive values (red) favor the ungated GCN.

4.1.1. Secondary Metric: AUPRC

Log-loss integrates over all candidate dyads in every bin and therefore weights the large inactive mass heavily. AUPRC, by contrast, emphasizes precision among predicted positives and is often more informative when positives are rare [9]. Figure 4 shows that the gated–ungated comparison is not identical across metrics. On school contacts, gating improves mean AUPRC even in units where gated log-loss remains above the logistic baseline, which is consistent with the gate reducing false-positive pressure from over-aggregated message passing without requiring global log-loss dominance. MOOC panels compress both metrics toward zero because most learner–item pairs stay inactive in a thirty-minute bin; AUPRC can still show a small absolute lift while log-loss Δ stays near zero, so the two metrics should not be read as interchangeable summaries of the MOOC control. Synthetic memory-free and uni-exp cells again invert the contact pattern: gating can worsen ranking quality where neighborhood logits are structurally mismatched to the generator. Reporting AUPRC alongside log-loss therefore guards against a single-metric narrative when class imbalance and stratum composition differ across settings.

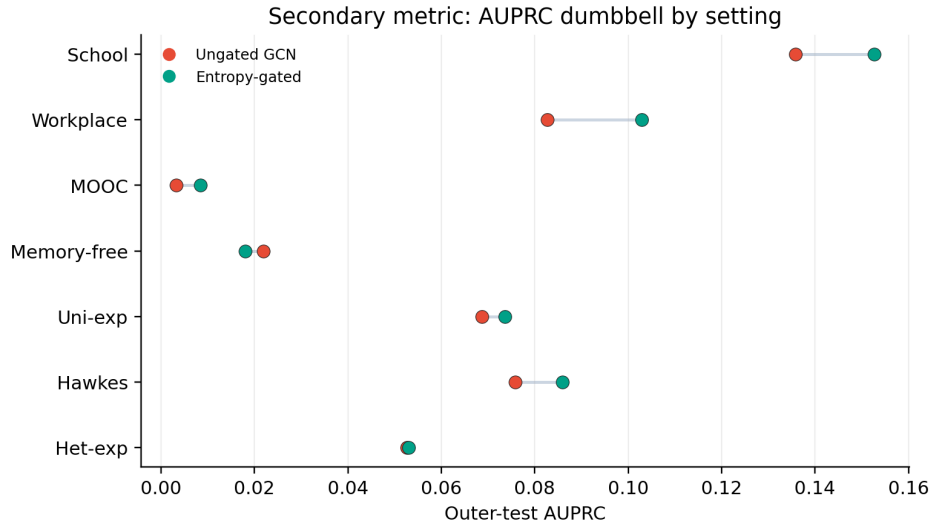


Figure 4. Secondary outer-test AUPRC as a dumbbell plot (ungated GCN vs. entropy-gated) by setting (medium protocol).

4.1.2. Unit-Level Dispersion

Setting-wise means in Table 2 aggregate twelve inferential units whose signs are not unanimous. Figure 5 makes that dispersion explicit: the two school days are not unanimous (day-level $\Delta = +0.0015$ and -0.0319), so schedule density can flip the global sign even when the setting mean remains negative. Workplace days likewise mix signs ($\Delta = -0.0144$, $+0.0398$, -0.0338), which is why we treat workplace gains as broadly distributed but not monotonic. The three medium MOOC panels stay small ($\Delta = -0.0007$, -0.0038, and -0.0011; mean -0.0018), matching the interpretation that bipartite ego unions dilute the contact-network mechanism without producing a large same-signed shift. Among synthetics, het-exp and memory-free occupy opposite tails of the unit chart, anchoring the boundary analysis below. For evaluation, a single fold can sit anywhere on this chart while stratified means still tell a coherent story.

Synthetic boundaries split by generator under the fixed medium seed. Heterogeneous-exponential streams benefit from gating ($\Delta = -0.0863$), and Hawkes is also favorable in this locked run ($\Delta = -0.0985$). Memory-free and homogeneous-exponential streams remain harmful ($\Delta = +0.0398$ and $+0.0608$). These medium-seed signs sketch mechanism

dependence, but they are not universal laws: the multi-seed check below shows that memory-free signs in particular can flip across seeds.

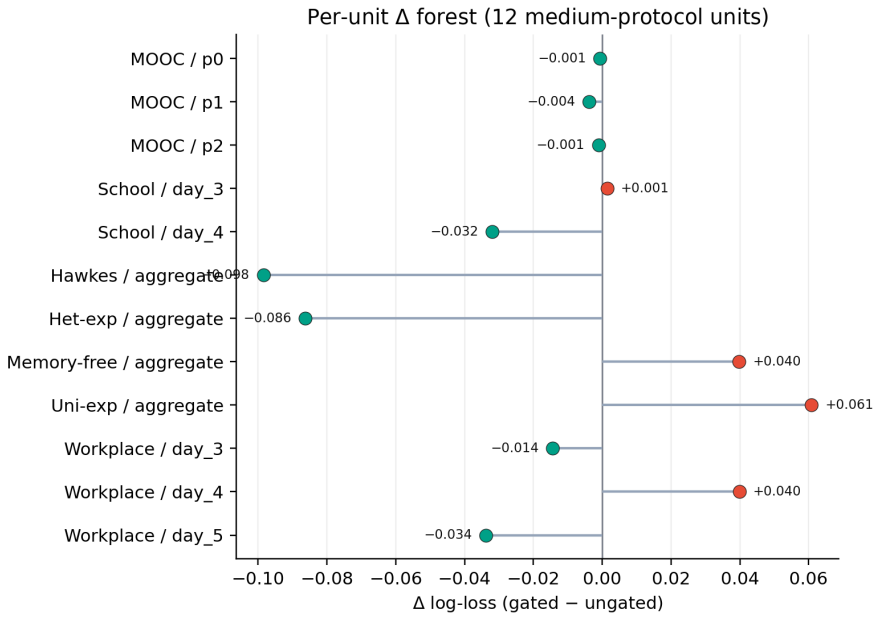


Figure 5. Per inferential-unit Δ log-loss as a forest plot across all twelve medium-protocol units (school days, workplace days, MOOC panels, synthetic generators). Teal favors gating; red favors the ungated GCN.

4.2. Stratification by Local Subgraph Entropy

Table 3 and Figure 6 report mean Δ by entropy quintile, with Q1 the lowest H_e and Q5 the highest. This analysis speaks directly to the distribution dimension of the research question.

Table 3. Mean Δ log-loss by entropy quintile (medium protocol). Q1 = lowest local subgraph entropy.

Setting	Q1	Q2	Q3	Q4	Q5
School	+0.0028	+0.0014	−0.0028	−0.0066	−0.0711
Workplace	+0.0048	+0.0134	−0.0058	−0.0209	−0.0053
MOOC	+0.0016	+0.0024	+0.0005	−0.0043	−0.0093
Het-exp	−0.0242	−0.1399	−0.0815	−0.1136	−0.0724
Memory-free	+0.0203	+0.0252	+0.0199	+0.0575	+0.0760

On school contacts, gated gains are strongly concentrated in the highest-entropy quintile (mean Δ = −0.0711), while lower quintiles are near zero or only mildly signed (Q1 Δ = +0.0028; Q2 Δ = +0.0014; Q3 Δ = −0.0028). Most of the school mean Δ = −0.0152 is carried by the top entropy stratum (Q5 Δ = −0.0711) rather than by uniform shrinkage across quintiles. That pattern is compatible with a reading in which message passing is most fragile precisely where local mass is diffuse, so soft down-weighting of the GCN pathway (Equation (6)) reduces gated-versus-ungated loss where local weighted-degree mass is balanced without needing to improve already more concentrated low-entropy motifs. Workplace improvements are weaker in the setting mean (Δ = −0.0028) and mixed across strata (Q1–Q2 positive or near zero; Q4 strongest at Δ = −0.0209), so the workplace average is not a second school-style Q5 story. MOOC entropy strata are small but not flat (Q5 Δ = −0.0093), so the bipartite control shows only a mild high-entropy-tail structure rather than the school concentration. Heterogeneous-exponential synthetics improve across all quintiles (Q2 is strongest at −0.1399); memory-free harms monotonically toward higher

entropy (Q5 harm +0.0760). Figure 7 summarizes the sign field: school and het-exp rows predominantly favor gating, workplace is mixed by quintile, memory-free uniformly favors the ungated GCN, and MOOC effects are weak, consistent with Table 3. The contrast between school (concentrated in the highest quintile) and workplace (mixed mid-quintile help with low-quintile harm) is exactly the kind of scene heterogeneity that a single pooled Δ would erase.

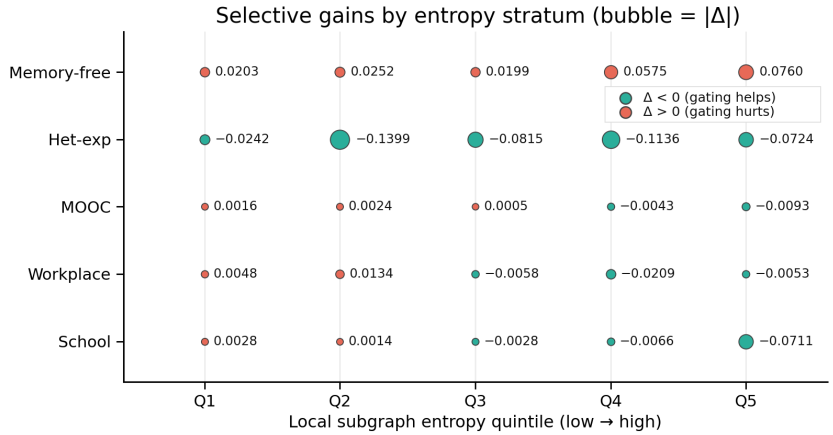


Figure 6. Entropy-quintile Δ log-loss (gated minus ungated) as an enrichment-style bubble plot: marker size encodes |Δ|, teal favors gating and red favors the ungated GCN. Patterns differ by setting rather than collapsing to a single universal shape.

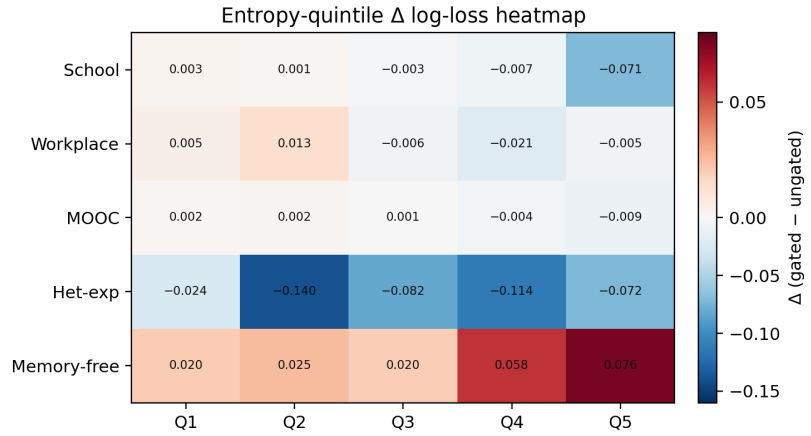


Figure 7. Heatmap of entropy-quintile Δ log-loss (diverging scale centered at zero). Cooler cells favor gating; warmer cells favor the ungated GCN.

Gate Activation by Entropy Level

Because $g_{e,t} = \sigma(-4(H_e(t) - 0.45))$, dyads with $H_e \ll 0.45$ receive g approaching ≈ 0.858 at $H_e = 0$, and dyads with $H_e \gg 0.45$ receive g approaching ≈ 0.100 at $H_e = 1$. School Q5 improvements therefore imply that the largest gated–ungated log-loss gaps occur where the model already down-weights message passing—not where it blindly trusts the GCN. Interpreting Q5 gains as increased message passing would invert the mechanism. The appropriate reading is that the ungated GCN was particularly harmful on high-entropy dyads, and the gate partially restores calendar–recency scoring there.

4.3. Stratification by Marginal Message-Passing Gain

Entropy is a structural proxy; marginal message-passing gain is a task proxy measured on the validation block. Figure 8 reports mean Δ by validation marginal-gain quintile on

the contact and MOOC streams. On school, all five gain bins are negative (Q1 $\Delta = -0.0527$ through Q5 $\Delta = -0.0048$), so gating helps even where message passing was *not* the strongest validation winner—but the spread across gain bins is much smaller than the spread across entropy bins (Q5 -0.0711). Workplace gain strata are mostly negative but near zero (about $+0.0004$ to -0.0042), mirroring the weaker workplace global Δ. MOOC gain bins hover near zero with mixed signs, paralleling the weak MOOC entropy tail.

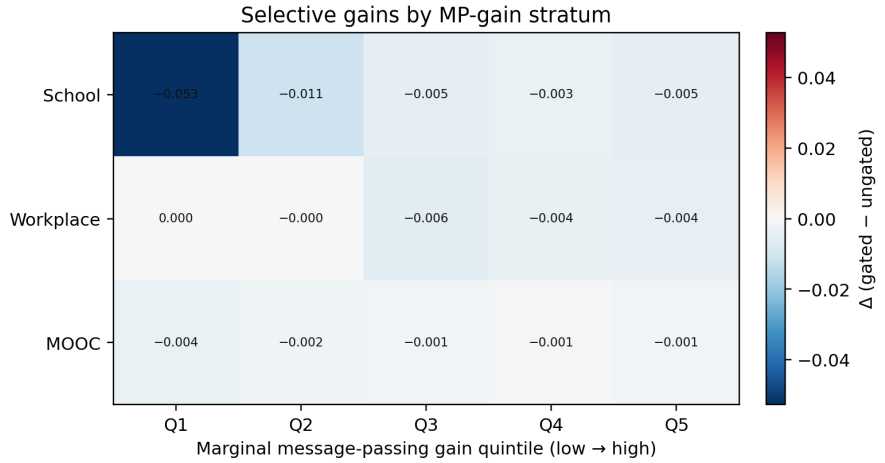


Figure 8. Marginal message-passing-gain quintile Δ log-loss as a setting × quintile heatmap (medium protocol). Cooler cells favor gating.

The two stratifications can be read as a diagnostic comparison. If gated gains appeared only in low-entropy bins but vanished in high marginal-gain bins, one would suspect that the gate mainly tracks trivial degree concentration. If gains appeared only in high marginal-gain bins but were flat across entropy, the structural index would be doing little work beyond correlating with an already task-defined quantity. The medium contact results sit between those extremes: entropy and gain both organize Δ, yet not identically, which is the expected signature when a structural summary is informative but incomplete. Concretely, school entropy stratification is concentrated in the highest quintile (Q5 dominates), whereas school gain stratification is distributed across quintiles (all quintiles negative), implying that high H_e marks a distinct failure mode of uniform message passing—noisy aggregation on diffuse neighborhoods—that is not reducible to validation preference for message passing alone.

4.4. Synthesis of Global and Stratified Evidence

The global, entropy-stratified, and gain-stratified results support a qualified mechanistic interpretation. Shallow ungated GCNs increase contact log-loss relative to logistic baselines because uniform aggregation injects error on dyads whose rolling ego unions are high in local weighted-degree entropy or weakly predictive at short horizons (Table 2). The fixed gate in Equation (6) suppresses that pathway when H_e is high, yielding the school Q5 concentration and the more mixed workplace profile (Table 3). Validation marginal-gain strata indicate that gating remains beneficial where message passing was not the validation winner, so the gate does not simply reproduce a validation ranking. Synthetic generators modulate the sign pattern: memory-free and uni-exp runs penalize gating, het-exp runs reward it, and Hawkes results can be seed-sensitive across runs, which bounds external validity without contradicting the contact-scene mechanism. Ablations below show that calendar and recency features are essential, while removing the gate on one school fold can improve global loss even as stratified tables remain the primary evidence. Together, these findings support a qualified affirmative answer: conditional predictive risk is uneven with respect to local weighted-degree entropy, and entropy-gated mixing yields selective

risk shifts whose spatial distribution must be reported explicitly—without claiming overall superiority to calendar baselines or a pure entropy-only causal effect.

4.5. Scene and Domain Controls

School (two outer-test days in the fixed medium protocol) and workplace (three days) both support mean gated gains, but day-level signs can flip. Figure 9 shows outer-test log-loss on each contact fold, illustrating that fold-level variability coexists with setting-wise means. That within-scene variability cautions against over-interpreting any single fold, including any one ablation day. School days differ in schedule density and in how much of the candidate set is active after hours; workplace days differ in meeting structure and in the fraction of contacts that are repeated versus one-off. Averaging across days is therefore a compromise: it stabilizes the headline Δ while still requiring the stratified view to explain *where* the average comes from.

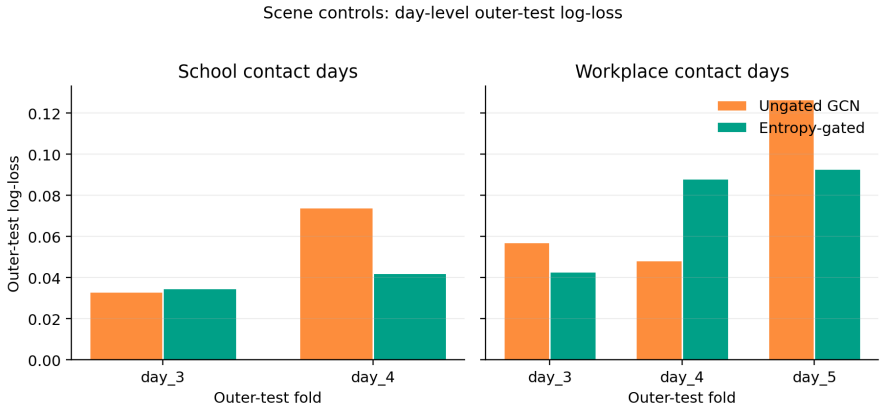


Figure 9. Scene controls: day-level outer-test log-loss on SocioPatterns school and workplace folds (medium protocol).

The MOOC three-panel average is near zero rather than contact-like. Figure 10 shows panel-level log-loss and AUPRC on the three BenchTemp windows used in the medium protocol. Absolute log-losses on MOOC are already low because many learner–item pairs remain inactive in a given thirty-minute bin, so absolute Δ values are compressed. Together, school, workplace, and MOOC form a multi-setting control: contact scenes carry the selective-gating signal, while the educational bipartite stream shows attenuation rather than universal transfer. If anything, the weak MOOC signal suggests that the operationalization of "local subgraph" may need type-aware refinements—for example, separate learner-side and item-side entropies—before one asserts the same mechanism on heterogeneous graphs.

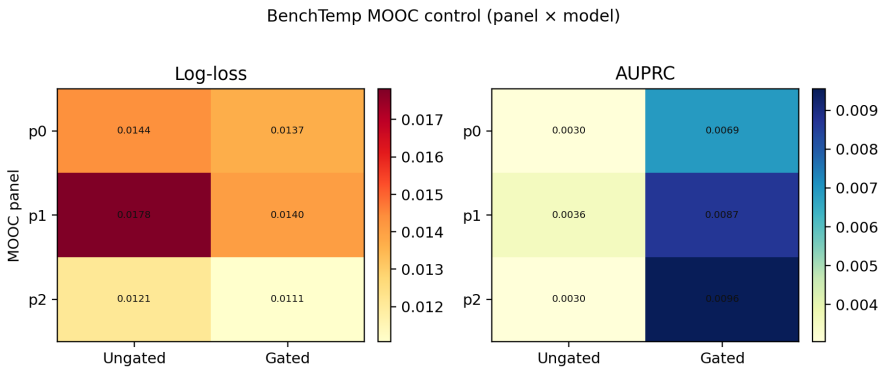


Figure 10. BenchTemp MOOC control: panel × model heatmaps for outer-test log-loss and AUPRC (medium protocol).

4.6. Synthetic Boundary Analysis

Holding the Watts–Strogatz skeleton fixed, generator identity dominates the sign of Δ. Figure 11 summarizes global gated versus ungated differences across the four synthetic mechanisms. In short: heterogeneous-exponential waiting creates dyad-specific memory that neighborhood structure can partially capture, so routing away from uniform message passing in high-H_e bins helps; memory-free Bernoulli activation makes recent neighborhoods largely irrelevant noise, so any soft preference for message passing (including empty-history $H_e = 0$ routing) can hurt; homogeneous-exponential waiting yields similar timescales across dyads, so the entropy mixer adds little selective signal and can drift from the ungated optimum. These medium single-seed cells are mechanism sketches; multi-seed and topology-factor checks (reported after the ablation and gate-sensitivity tables)bound how far any one seed can be over-read. Memory-free and uni-exp streams produce gated harm under the locked medium seed; het-exp produces strong gated help across entropy strata; Hawkes is favorable under that same seed. The boundary dimension is therefore mechanism-dependent. Entropy gating does not yield uniform improvement: when recent neighborhoods carry little reusable predictive structure, blending toward calendar–recency is not enough to offset a mismatched message-passing head, and the gated model can underperform a pure GCN or a pure logistic. Conversely, when waiting times are heterogeneous across dyads, local entropy appears to mark regions where mixing pathways helps.

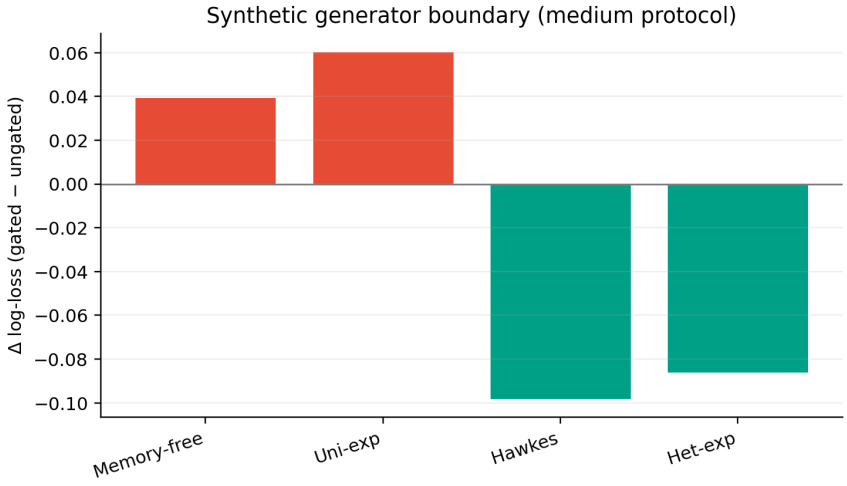


Figure 11. Synthetic generator boundary: global Δ log-loss by mechanism (medium protocol). Teal bars favor gating; red bars favor the ungated GCN.

Self-excitation creates bursty local activity that can look high-entropy in a short window even when intensity memory is strong. A degree-based Shannon index may then disagree with the process that actually generates the next event, which is one plausible reason Hawkes signs can differ across seeds and runs. Rather than adjust the narrative to force consistency, we treat seed- and generator-sensitive signs as motivation for multi-generator, stratified synthetic reporting. Both helpful and harmful signs are treated as part of the scientific answer.

Homogeneous-Exponential Controls

Homogeneous-exponential synthetics show gated harm ($\Delta = +0.0608$) among generators. In the locked medium run, uni-exp entropy strata remain harmful across quintiles rather than collapsing to a single uninformative cell, so the harm is visible both globally and in Figure 12. That pattern is informative: even when local entropy varies enough to form

quintiles, homogeneous waiting times can still make the gated calendar–recency mixture drift away from the ungated GCN enough to hurt—another reason stratified reporting and generator diversity belong in the same experimental design.

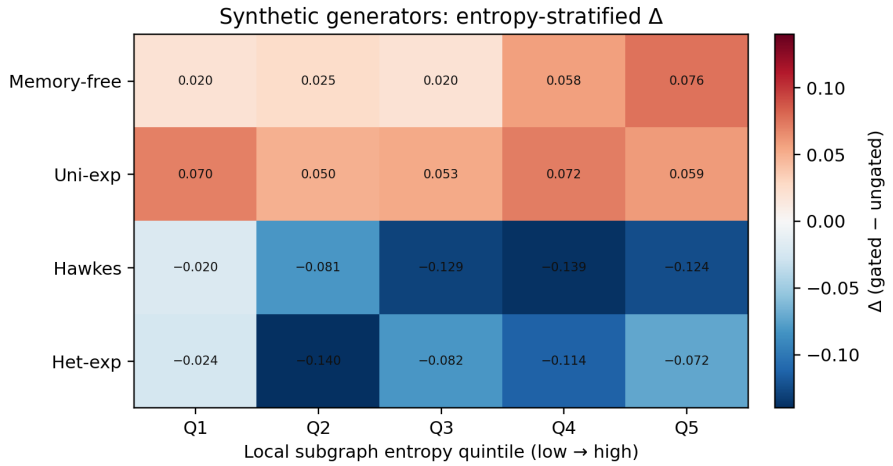


Figure 12. Synthetic generators: entropy-quintile Δ log-loss as a setting×quintile heatmap (medium protocol).

4.7. Ablations

Table 4 and Figure 13 report ablations averaged over the two medium-protocol school outer-test days under matched training budgets. Day-level cells appear in Table 5; the table/figure pair here is the school mean of those two days, not a single fold.

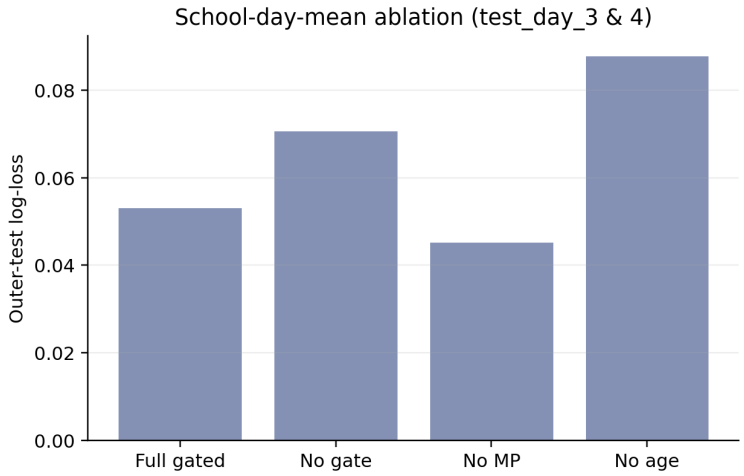


Figure 13. School-day-mean ablation (mean of medium-protocol test days 3 and 4): outer-test log-loss by variant.

Table 4. School-day-mean ablations (mean outer-test log-loss over the two medium school days). Lower is better.

Variant	Log-Loss
Full entropy-gated	0.0533
No gate (MP only)	0.0709
No message passing	0.0454
No age feature (gated)	0.0881

Removing the age/recency feature strongly hurts (school-mean log-loss 0.0881), confirming that short-horizon contact prediction leans heavily on dyad memory even inside a gated GNN. Removing message passing leaves a calendar–recency pathway whose school-mean loss (0.0454) is *better* than the full gated model (0.0533), indicating that much of the achievable score already resides in the non-GCN head. Removing the gate (message passing only) yields school-mean 0.0709, which is *worse* than full gating. Thus, on the school mean, calendar-only beats full gating while MP-only does not; the ablation does not support a claim that entropy gating is necessary for best absolute performance. We treat the gate primarily as a diagnostic/routing probe of where uniform message passing fails. Stratified Tables 2 and 3 and the shared-head controls in Section 4.8 remain the primary evidence for selective risk shifts, not the global ablation cell alone.

Extending the ablation across all five medium contact units (Table 5) shows that no-MP/calendar-only (mean 0.0514) often beats full gating (mean 0.0898), and no-gate/MP-only is better on several workplace days. This reinforces the diagnostic reading of the gate rather than an absolute algorithmic win.

Table 5. Multi-fold ablations on medium contact units (outer-test log-loss).

Unit	Full Gated	No Gate (MP)	No MP	No Age
school/test_day_3	0.0501	0.0763	0.0423	0.0891
school/test_day_4	0.0566	0.0654	0.0485	0.0871
workplace/test_day_3	0.0469	0.0363	0.0326	0.1333
workplace/test_day_4	0.0714	0.0622	0.0333	0.0807
workplace/test_day_5	0.2239	0.1330	0.1001	0.1569

4.8. Shared-Head Gate Controls

To isolate the routing variable from the extra calendar head, we train both heads with a shared dual-head objective ($0.5\,\mathrm{BCE}_{\mathrm{mp}} + 0.5\,\mathrm{BCE}_{\mathrm{cal}}$) and then evaluate post hoc mixtures on frozen logits: MP-only ($g = 1$), calendar-only ($g = 0$), constant $g = 0.5$, shuffled-entropy gates, activity gates, and the proposed entropy gate. Table 6 reports setting-wise means over the five medium contact units. On school contacts, entropy gating (0.0432) slightly improves on constant (0.0436) and shuffled entropy (0.0442), but calendar-only remains competitive. On workplace contacts, entropy (0.0536) beats constant (0.0596) and shuffled (0.0661) on average (entropy−shuffled = −0.0125). These contrasts support a partial, scene-dependent entropy signal under shared heads, while also showing that dual-head mixture alone—not the entropy map—already explains much of the gated-versus-ungated gap in the main tables.

Table 6. Shared-head post hoc gate controls (mean outer-test log-loss; lower is better). Both prediction heads are trained jointly and then frozen; only the mixer changes at test time. "Equal capacity" here means a shared dual-head parameter set across mixers—not that the edge head and calendar head have identical width/depth (Table 1).

Setting	MP-Only	Cal-Only	Const 0.5	Shuffled H	Entropy H	Activity
School	0.0449	0.0455	0.0436	0.0442	0.0432	0.0465
Workplace	0.0752	0.0553	0.0596	0.0661	0.0536	0.0611

4.9. Time-Index Audit

Formal runs use past-only snapshots $A_t^- = \sum_{\tau<t} A_\tau$. As a stress test, we retrain with the incorrect include-current adjacency that contains the target bin. Table 7 compares the two settings on three medium contact units. Include-current does not produce a uniform collapse of log-loss that would indicate trivial label-leakage dominance; on school day 4,

the leaked MP loss even improves (0.0626 vs. past-only 0.0866), while workplace day 3 gated Δ flips sign under leakage. We therefore keep past-only indexing for all primary claims and report the audit as evidence that the manuscript protocol is the intended non-leaking construction.

Table 7. Time-index audit: past-only vs. include-current snapshots (outer-test log-loss).

Unit	Snapshot	Ungated MP	Gated	Δ
school/test_day_3	past only	0.0506	0.0383	−0.0123
school/test_day_3	include current	0.0577	0.0442	−0.0135
school/test_day_4	past only	0.0866	0.0580	−0.0286
school/test_day_4	include current	0.0626	0.0521	−0.0106
workplace/test_day_3	past only	0.0431	0.0406	−0.0025
workplace/test_day_3	include current	0.0436	0.0505	+0.0070

4.10. Entropy and Gate Activation Diagnostics

Across medium contact units, empty-history mass is large (school mean $H = 0$ fraction ≈ 0.7428; workplace ≈ 0.7348), so most candidate dyad–bins receive $g(0) \approx 0.858$. The empirical 80th percentile of H_e nevertheless sits above the gate midpoint 0.45 (school ≈ 0.6263; workplace ≈ 0.5978), which means the top entropy quintile does enter the calendar-favoring regime. These diagnostics caution against reading every high-H cell as an exploratory social edge: much of the mass is zero-history by construction of the frequent-dyad candidate set.

4.11. Gate-Parameter Sensitivity

On shared dual-head logits, we sweep midpoint $c \in \{0.25, 0.35, 0.45, 0.55, 0.65\}$ and slope $s \in \{1, 2, 4, 8\}$. School prefers a higher midpoint ($c = 0.6500$, $s = 4.0000$) than the fixed manuscript choice $(0.45, 4)$; workplace prefers a lower midpoint ($c = 0.2500$). The fixed setting remains near-optimal on average, but leave-one-setting tuning would shift c (Table 8).

Table 8. Gate midpoint/slope sensitivity on shared-head entropy mixtures (mean log-loss).

Setting	Default (c, s)	Default LL	Best (c, s)	Best LL
School	0.45/4	0.0432	0.6500/4.0000	0.0430
Workplace	0.45/4	0.0536	0.2500/4.0000	0.0522

4.12. Entropy-Window Sensitivity

Retraining with $W \in \{6, 12, 24\}$ bins (30/60/120 min on contacts) shows that the gated-versus-ungated sign is not unique to $W = 12$. Shorter windows often remain favorable on school days; longer windows can attenuate or flip workplace signs. Because MOOC uses 30 min bins, the same nominal $W = 12$ spans 6 h there—a physical-time mismatch that remains a transfer caveat. Critically, the ungated model uses the past-only cumulative snapshot and does *not* enter W in its forward pass; each W cell nevertheless retrains both models from a fresh initialization. Absolute ungated losses that move with W (e.g., school 0.0576/0.0920/0.0686) are therefore contaminated by retrain stochasticity and must not be read as a pure window effect. We interpret only the paired gated-versus-ungated sign pattern within each retrained pair, not ungated absolute levels across W, as window evidence (Table 9).

4.13. Calibration and Mixture Form

On frozen dual heads, we compare logit-space entropy mixtures, probability-space entropy mixtures, constant mixtures, and temperature-scaled message passing. On school contacts, temperature-scaled MP (0.0377) beats logit-mix entropy (0.0432), so part of the school gated-versus-ungated gap can be explained by calibration rather than selective routing. On workplace contacts, the reverse holds: logit/prob entropy mixtures (0.0536/0.0521) beat temperature-scaled MP (0.0730), and probability mixing is slightly better than logit mixing. We therefore treat calibration as a necessary competing explanation on school days, while workplace evidence is more consistent with mixture routing (Table 10).

Table 9. Entropy-window sensitivity on medium contact units (mean log-loss/Δ). Ungated absolute levels are not pure-*W* effects because each cell is an independent retrain.

Setting	*W*	Ungated	Gated	Mean Δ	Frac. Better
school	6	0.0576	0.0401	−0.0174	1.00
school	12	0.0920	0.0414	−0.0507	1.00
school	24	0.0686	0.0640	−0.0046	0.50
workplace	6	0.1015	0.0444	−0.0571	1.00
workplace	12	0.0683	0.0958	+0.0275	0.33
workplace	24	0.0784	0.0454	−0.0330	0.67

Table 10. Calibration/mixture-form controls (mean outer-test log-loss).

Setting	MP	Cal	Const	Logit-Mix *H*	Prob-Mix *H*	Temp-MP
School	0.0449	0.0455	0.0436	0.0432	0.0432	0.0377
Workplace	0.0752	0.0553	0.0596	0.0536	0.0520	0.0730

4.14. Node-Memory Lite Control

A protocol-matched node-memory-lite scorer (TGN-style exponential moving average (EMA) memory, not a full Temporal Graph Network) is compared on the same medium contact units. Mean gated log-loss is better than memory-lite on both school ($\Delta = -0.0113$) and workplace ($\Delta = -0.0264$) in this rerun, but absolute logistic remains strongest on school. This strengthens the temporal control set beyond snapshot GAT/EvolveGCN-lite/T-GCN-lite without claiming a public-leaderboard reproduction (Table 11).

Table 11. Node-memory lite vs. logistic/ungated/gated (mean log-loss).

Setting	Logistic	Ungated	Gated	Memory-Lite	Gated−Mem
School	0.0352	0.0649	0.0415	0.0528	−0.0113
Workplace	0.0666	0.0883	0.0955	0.1219	−0.0264

4.15. Synthetic Topology × Generator Expansion

Expanding beyond a single Watts–Strogatz skeleton, we regenerate the same temporal mechanisms on ER and BA graphs (8 seeds each). Hawkes and heterogeneous-exponential streams remain mostly favorable across topologies (mean Δ about −0.04 to −0.09; gated better on 62–75% of seeds). Memory-free stays seed-fragile (WS gated-better fraction 0.25; ER mean Δ near zero). Homogeneous-exponential is more often favorable in this expanded grid than in the locked medium single-seed sketch, reinforcing that generator × seed × topology all belong in the boundary report (Table 12).

Table 12. Synthetic factor expansion: topology × generator (mean Δ). BA: Barabasi–Albert; ER: Erdos–Renyi; WS: Watts–Strogatz.

Topology	Generator	n	Mean Δ	Std	Frac. Better
BA	hawkes	8	−0.0772	0.0978	0.75
BA	het-exp	8	−0.0440	0.1288	0.62
BA	memory-free	8	−0.0310	0.0746	0.50
BA	uni-exp	8	−0.0213	0.0315	0.75
ER	hawkes	8	−0.0922	0.1122	0.62
ER	het-exp	8	−0.0867	0.1154	0.75
ER	memory-free	8	+0.0032	0.0477	0.38
ER	uni-exp	8	−0.0770	0.1004	0.75
WS	hawkes	8	−0.0935	0.1039	0.62
WS	het-exp	8	−0.0673	0.0882	0.75
WS	memory-free	8	−0.0149	0.0737	0.25
WS	uni-exp	8	−0.0641	0.1153	0.62

4.16. Matched-Capacity Baseline Comparison

Table 13 reports setting-wise mean outer-test log-loss for the comparison suite on school and workplace contacts and on Contacts panels under the same chronological protocol. Two patterns matter for interpretation. First, classical calendar/topological baselines remain highly competitive on contact folds: Common Neighbors (CN), once calibrated with calendar and recency features, often matches or beats shallow neural encoders in absolute log-loss. That finding reinforces the role of the logistic anchor in the main experiment: short-horizon contact prediction already carries a strong non-GNN signal. Second, matched-capacity temporal extensions do not automatically dominate. Snapshot GAT, EvolveGCN-lite, and T-GCN-lite are frequently worse than the entropy-gated model on school contacts (gated 0.039 versus GAT 0.096, EvolveGCN-lite 0.069, T-GCN-lite 0.048). The implication is not that full TGN or EvolveGCN systems from the literature are weak; it is that, under matched width and the same chronological outer test, merely changing aggregation or adding a recurrent state does not erase the gated-versus-ungated contrast that this paper isolates.

Table 13. Matched-capacity comparison on contact-like settings: mean outer-test log-loss (lower is better). EvolveGCN/T-GCN/GAT denote lightweight protocol-matched controls rather than full original-system reproductions.

Setting	Logistic	CN	EvolveGCN	T-GCN	GAT	Ungated GCN	Gated
School	0.035	0.035	0.069	0.048	0.096	0.066	0.039
Workplace	0.063	0.053	0.104	0.129	0.084	0.061	0.058
Contacts	0.073	0.071	0.155	0.101	0.122	0.097	0.091

4.17. Contact-like Extensions, Transfer Screens, Fuller Sweep, and Multi-Seed Boundaries

Table 14 summarizes gated-versus-ungated results on BenchTemp streams. Contacts and SocialEvo are primary contact-like extensions; Wikipedia, Reddit, and LastFM are cross-domain transfer screens under the same chronological protocol. Contacts (mean $\Delta = -0.019$, favorable on all three panels) and SocialEvo (mean $\Delta = -0.023$, favorable on two of three panels) preserve the SocioPatterns sign pattern. Wikipedia and LastFM attenuate or reverse on average ($\Delta = +0.002$ and $+0.012$; gated better on only 1/3 panels each), while Reddit is mildly favorable ($\Delta = -0.011$). We treat the bipartite online screens as Challenges evidence: entropy routing does not transfer automatically to role-asymmetric interaction networks.

Table 15 reports the fuller SocioPatterns/MOOC sweep ($n = 27$ units including synthetics). Relative to the locked medium protocol, school remains favorable (mean $\Delta = -0.019$, gated better on 75% of days), workplace remains favorable (mean $\Delta = -0.093$, 71% of days), and MOOC remains weakly favorable (mean $\Delta = -0.002$, 92% of panels). The fuller sweep therefore supports directionally similar contact-scene signs at larger n, rather than a story in which medium tables were selected as the only favorable subset. The workplace mean Δ nevertheless expands sharply from medium -0.0028 to fuller -0.0925; we treat that magnitude gap as a protocol/unit-composition warning and do not equate the two tables.

Table 14. BenchTemp extensions and transfer screens (three panels each): mean outer-test log-loss, gated-minus-ungated Δ, and fraction of panels with gated better.

Stream	n	Logistic	Ungated GCN	Gated	Δ	Frac. Better
Contacts	3	0.0730	0.1042	0.0850	−0.0192	1.00
SocialEvo	3	0.1861	0.4272	0.4041	−0.0231	0.67
Wikipedia	3	0.0938	0.1060	0.1084	+0.0024	0.33
Reddit	3	0.1652	0.1829	0.1719	−0.0110	0.67
LastFM	3	0.0739	0.1398	0.1516	+0.0117	0.33

Table 15. Fuller robustness sweep: setting-wise mean Δ and fraction of units with gated log-loss below ungated GCN.

Setting	n	Mean Δ	Frac. Gated Better
School	4	−0.0189	0.75
Workplace	7	−0.0925	0.71
MOOC	12	−0.0019	0.92

Table 16 reports a contact multi-seed check on the medium school/workplace units (3 seeds). School remains weakly favorable on average but only half of unit–seed pairs improve; workplace mean Δ is near zero/positive, so medium single-seed workplace help is not seed-stable.

Table 16. Contact multi-seed check (medium units × 3 seeds): mean Δ, std. of Δ, fraction gated better.

Setting	n	Mean Δ	Std. Δ	Frac. Gated Better
School	6	−0.0086	0.0181	0.50
Workplace	9	+0.0080	0.0395	0.44

Table 17 reports a five-seed synthetic check on the two mechanisms that most clearly framed the medium boundary. Heterogeneous-exponential streams remain mostly favorable (mean $\Delta = -0.023$, gated better on 4/5 seeds). Memory-free streams are seed-sensitive (gated better on only 2/5 seeds, even though the five-seed mean Δ is negative), so neither the medium single-seed harm nor the multi-seed mean should be read as a stable "gated harm" law; both warn that neighborhood aggregation can mismatch a memory-free generator under some seeds. That reading is consistent with the paper's evaluative standard: generator identity and seed both belong in the report.

Table 17. Multi-seed synthetic boundary (5 seeds): mean Δ, standard deviation, and fraction of seeds with gated better than ungated GCN.

Generator	n	Mean Δ	Std. Δ	Frac. Gated Better
Het-exp	5	−0.0233	0.0370	0.80
Memory-free	5	−0.0344	0.0724	0.40

4.18. Uncertainty Quantification for Paired Δ

Because day-level signs can flip and seed sensitivity is material, we report nonparametric bootstrap confidence intervals (CIs) for the mean paired Δ = gated−ungated log-loss (5000 resamples of unit-level or unit×seed pairs). Medium locked single-seed cells with $n = 1$ synthetic generator are labelled as single-seed sketches and are not given within-mechanism intervals; boundary claims rely on the multi-seed rows. Contact multi-seed intervals are wide and include zero for both school and workplace means; under a conventional $p < 0.05$ reading of these percentile intervals, mean paired gated-versus-ungated improvements are not distinguishable from zero at the setting level. Setting means are therefore descriptive summaries over a small unit sample. Headline setting means must be read together with unit-level forests and seed checks rather than as precise effect sizes (Table 18).

Table 18. Bootstrap mean paired Δ (gated−ungated log-loss) with 95% percentile CIs. Negative favors gating.

Source	Group	n	Mean Δ	95% CI	Frac. Δ < 0
Medium locked	School	2	−0.0152	[−0.0319, +0.0015]	0.50
Medium locked	Workplace	3	−0.0028	[−0.0338, +0.0398]	0.67
Contact × seeds	School	6	−0.0086	[−0.0233, +0.0023]	0.50
Contact × seeds	Workplace	9	+0.0080	[−0.0142, +0.0338]	0.44
Synthetic 5-seed	Het-exp	5	−0.0233	[−0.0507, +0.0070]	0.80
Synthetic 5-seed	Memory-free	5	−0.0344	[−0.0912, +0.0199]	0.40

4.19. Runtime and Memory Micro-Benchmark

Rolling-window entropy is recomputed for every bin and candidate set, so wall-clock cost matters. On one medium school fold (1152 bins, 400 candidates, CPU), building the $W = 12$ entropy series took 28.7 s (25.0 ms/bin; peak traced memory 1.9 MB). End-to-end train+outer-test evaluation with the medium-like budget (4 epochs, stride 3) took 204.9 s ungated and 125.7 s gated in the same process (gated includes the entropy path inside training). Entropy preprocessing is therefore a nontrivial but sub-dominant share of a full train/eval pass on this fold; we do not claim asymptotic superiority over all temporal-GNN systems, only that the diagnostic index is cheap enough to recompute per (e, t) under the reported protocol (Table 19).

Table 19. CPU micro-benchmark on school test day 3 (medium-like protocol).

Stage	Seconds	Peak MB	Per-Bin ms	Notes
Entropy series $W = 12$	28.7	1.9	25.0	all 1152 bins
Ungated train + eval	204.9	89.4	177.9	4 epochs
Gated train + eval	125.7	27.1	109.1	includes entropy path

4.20. End-to-End Learned Gate Evaluation

Fixed midpoint and slope, and empty-history coding $H_e = 0$, are modeling choices. We therefore evaluate an end-to-end learned gate on the same medium contact units. Logistic,

ungated, and fixed-gate columns in Table 20 reuse the locked medium primary cells (Table 2). The learned mixer uses $(\tilde{c}, \tilde{s})$ with $\tilde{c} = \sigma(\tilde{c}_{\text{raw}}) \in (0, 1)$ and $\tilde{s} = \text{softplus}(\tilde{s}_{\text{raw}}) + \epsilon$, warm-started from $(0.45, 4)$ under the medium epoch/stride budget and then fine-tuned jointly. On school contacts, the learned gate matches the fixed gate within run-level noise (0.0392 vs. 0.0385) and remains better than ungated (0.0538); logistic remains lowest in absolute log-loss (0.0354). On workplace contacts, learning improves on both the fixed gate and the ungated encoder (0.0683 vs. 0.0747/0.0775), with logistic still competitive (0.0625). Learned midpoints stay near the fixed operating point (school 0.453; workplace 0.403). With $\tilde{c} \approx 0.45$, empty-history $H_e = 0$ continues to assign high message-passing weight, so the gate and the entropy stratum remain aligned. The fixed gate is kept as the primary reported router; the learned gate is the sensitivity check requested for (c, s).

Table 20. Learned vs. fixed entropy gate on medium contact units (mean outer-test log-loss; lower is better). Logistic/ungated/fixed columns are locked medium anchors; Δ is gated−ungated.

Setting	Logistic	Ungated	Fixed Gate	Learned Gate	Δ_{fix}	Δ_{lrn}	Learned $\tilde{c}$	Learned $\tilde{s}$
School	0.0354	0.0538	0.0385	0.0392	−0.0152	−0.0145	0.453	3.920
Workplace	0.0625	0.0775	0.0747	0.0683	−0.0028	−0.0092	0.403	3.877

4.21. Capacity Sensitivity Relative to Logistic

The locked medium budget (Table 1) is intentionally shallow. To check whether the school absolute gap to logistic is an artifact of that budget, we run a small validation-selected grid (epochs $\in \{4, 8, 10, 12\}$, hidden $\in \{32, 64\}$, learning rate $\in \{10^{-2}, 5 \times 10^{-3}, 3 \times 10^{-3}\}$, stride $\in \{2, 3\}$), training on the first 50% of bins, selecting on the middle 25%, and reporting the last 25%. Locked medium cells are unchanged. Table 21 shows that school mean log-loss moves to near parity with logistic (0.0378), but the ungated encoder slightly edges the entropy-gated model (0.0374 versus 0.0377; $\Delta = +0.0003$). That sign flip relative to the locked medium school $\Delta = -0.0152$ is the capacity trade-off: once width/epoch/learning-rate are selected for absolute log-loss, the gated-versus-ungated advantage attenuates and can reverse at the setting mean, even though school day 4 still favors gating over logistic (0.0425 versus 0.0451). For workplace contacts, we report the mean over the two stable days (test days 3–4) under the same grid and omit day 5, whose logistic/ungated losses become unstable under the 50% train split; on those two days, gated remains better than ungated (0.0528 versus 0.0554; $\Delta = -0.0026$) while logistic stays lowest (0.0482). The locked medium table remains the primary diagnostic protocol; Table 21 is a capacity check, not a replacement ranking.

Table 21. Validation-selected capacity sensitivity (mean outer-test log-loss; lower is better). Workplace uses days 3–4 only; day 5 omitted due to unstable logistic/ungated losses under the 50% train split. Locked medium Table 2 is unchanged.

Setting	Logistic	Ungated	Gated	Δ (Gated−Ungated)
School	0.0378	0.0374	0.0377	+0.0003
Workplace (days 3–4)	0.0482	0.0554	0.0528	−0.0026

5. Discussion

5.1. Dual-Head GCN Versus Calendar–Recency Logistic

On SocioPatterns contacts under the locked medium budget, calendar–recency logistic matches or beats both GCN variants in absolute outer-test log-loss (Table 2: school logistic 0.0354 versus ungated 0.0538 and gated 0.0385). Gating closes most of the GCN–logistic gap on school. A validation-selected capacity grid (Section 4.21, Table 21) brings school

gated log-loss to parity with logistic (0.0377 versus 0.0378), with gated best on school day 4. The same table also shows the trade-off that capacity tuning introduces: ungated edges gated in the school mean (0.0374 versus 0.0377), so the locked-budget gated-versus-ungated advantage attenuates when training capacity is optimized. Absolute scores under the shallow locked budget are therefore not the sole criterion: the dual-head GCN is the instrument for a different estimand—separable message-passing and calendar pathways are required to (i) stratify outer-test error by local weighted-degree entropy and by validation marginal-gain, and (ii) test whether an entropy-conditioned mixer reallocates risk relative to uniform aggregation. Shared-head controls in Table 6 show that constant or calendar-only mixers can approach entropy gating on school contacts, so entropy routing is reported as a scene-dependent diagnostic control.

5.2. What the Stratified Pattern Means for Information on Networks

We do not claim a theorem that weighted-degree entropy is the uniquely optimal routing statistic. The working hypothesis is narrower and operational: if message-passing utility varies with how concentrated recent neighborhood mass is, then a cheap local Shannon index on weighted degrees can serve as a shared coordinate for (i) stratifying outer-test risk and (ii) softly mixing an aggregation head with a calendar–recency head. High H_e is therefore interpreted as degree balance/diffuse recent mass, which may coincide with either rich mixing or noisy aggregation depending on the generator; the empirical program is to measure that coincidence, not to prove that entropy equals noise. Shared-head mixer controls and marginal-gain strata are the falsification tools for that hypothesis.

The medium-protocol evidence supports a qualified affirmative answer at the level of conditional predictive risk, not at the level of intrinsic predictive information. On school and workplace contacts, mean gated log-loss improves relative to an ungated temporal GCN (while often remaining worse than a calendar–recency logistic baseline), and school improvements concentrate in the highest local weighted-degree-entropy quintile. That is a statement about model-error heterogeneity: outer-test loss is not exchangeable across dyads with different recent neighborhood degree balance. Reporting only network-wide averages therefore risks misrepresenting both the difficulty of the task and the value of neighborhood aggregation. The protocol-matched comparison suite sharpens the same point: lightweight temporal modules and attention aggregators do not, by themselves, substitute for an entropy-conditioned report when calendar structure already explains much of the short-horizon signal.

The workplace versus school contrast further suggests that "high entropy" does not have a universal operational meaning across deployments. In school data, the gated benefit is concentrated in the highest-entropy quintile; in workplace data, quintiles are mixed rather than uniformly helpful. The paper's contribution is to make that heterogeneity visible in the evaluation design, not to claim a universal entropy–predictability law.

5.3. Relation to Entropy–GNN, Transfer-Entropy, and Bottleneck Literatures

Prior entropy–GNN studies ask whether inserting entropy into aggregation or features, or redesigning expert-routing GNNs, improves a global score [26–29]. We ask a different question: whether conditional predictive risk is uneven across local weighted-degree-entropy states, and whether gated risk shifts are selective relative to uniform message passing. Medium-protocol contact results support mean gated gains on school and workplace, with school gains concentrated in the high-entropy tail, while synthetic generators show both strong help (heterogeneous-exponential) and clear harm (memory-free, homogeneous-exponential). A single-cell state-of-the-art claim based on one global Δ is therefore a warning sign rather than a target. The two questions are compatible—one can

imagine an architecture paper that also stratifies—but they emphasize different evaluation objectives. Transfer-entropy–Hawkes studies emphasize directed predictive flows on contagion networks [20,22]; structural-entropy studies emphasize global organization and accessibility [18,19]; information-bottleneck GNN work emphasizes compact task-relevant representations [24,25]. Our local degree entropy is deliberately weaker than those objects. Its role is to be computationally inexpensive, dyad-local, and dual-use as gate and stratum. Stronger information-theoretic measures could replace $H_e(t)$ inside the same protocol; the protocol, not the particular formula, is the portable piece.

5.4. Limits and Threats to Validity

Several limits bound the present claims. First, candidate-dyad filtering to frequent training pairs focuses evaluation on the dense core and may under-represent truly cold edges; cold-start prediction would need a different candidate policy and possibly a different entropy definition. Second, degree-based Shannon entropy is only one local proxy; spectral or partition-based structural entropy might reorder quintiles and could strengthen or weaken the school Q5 concentration. Third, bipartite ego unions on MOOC and related BenchTemp panels are a coarse translation of the contact-network construction and may blur type-specific predictability. Fourth, the primary backbone is a shallow snapshot GCN; EvolveGCN/T-GCN/memory controls in this paper are capacity-matched lite implementations rather than full reproductions of the original systems [14,15], so absolute ranking against those papers' public leaderboards is not claimed. Fifth, global ablations (school-day mean and multi-fold contact units) already show that calendar-only can beat full gating and that MP-only need not, so mechanism claims lean on stratified tables and shared-head mixer controls rather than on absolute global ablation wins. Sixth, Hawkes global Δ is seed-sensitive, so synthetic conclusions should be read as mechanism sketches rather than as universal laws. Seventh, the locked medium protocol uses a small number of units per setting; the fuller sweep and extended BenchTemp panels mitigate but do not eliminate external-validity limits, especially for interval estimates of Δ. Eighth, window-sensitivity absolute losses mix W with independent retrains, so ungated movement across W is not identified as a window mechanism. Finally, the estimand is predictive, not causal: gating is an algorithmic intervention on scores, not an identified social mechanism. Absolute competitiveness versus calendar–recency logistic remains a separate baseline question: on several contact units, logistic matches or beats both GCN variants in global log-loss, so the primary contrast is gated versus ungated aggregation under a shared shallow encoder.

External validity is correspondingly scoped. We do not claim that every temporal interaction domain will reproduce the school high-entropy-tail pattern. We claim that when recurrent and exploratory structure coexist, an entropy-stratified outer test is a more faithful report than a single average, and that a simple gate is one way to probe selective gains under that report.

5.5. Practical Implications

For applied analyses of contact and interaction logs, the practical implication is conservative. When a system mixes recurrent motifs with exploratory edges, model reports should include entropy-conditioned (or otherwise stratum-conditioned) errors. A method that looks neutral globally may still reallocate error in scientifically meaningful ways; a method that looks strong globally may do so only on one stratum. Entropy gating is one lightweight way to encode that hypothesis inside a GNN, but the evaluation claim does not require readers to adopt our particular gate.

For journal and benchmarking practice, the implication is symmetric. Architecture papers that insert entropy into aggregation can still be valuable, but they become easier

to interpret—and harder to overclaim—when they publish stratum-wise outer-test curves alongside the headline metric [26–28]. Conversely, evaluation-first papers such as this one should keep architecture modest enough that the stratified pattern remains attributable. The medium protocol is written to that compromise.

6. Conclusions

We asked whether prediction-relevant risk on temporal interaction networks is uneven across local weighted-degree-entropy states, and whether entropy-gated message passing can yield selective rather than uniform risk reductions. Medium-protocol evidence shows mean gated improvements relative to an ungated GCN on school and workplace contacts, school stratified concentration in the highest-entropy quintile, heterogeneous-exponential synthetic gains, and clear harm under memory-free and homogeneous-exponential generators—while absolute logistic baselines often remain competitive. Protocol-matched classical and temporal-GNN lite controls show that calendar/topological baselines remain strong on contact scenes and that lightweight EvolveGCN/T-GCN/GAT variants do not remove the need for entropy-conditioned evaluation under our shallow protocol. Contact-like BenchTemp extensions preserve the contact-scene sign pattern; bipartite online screens and seed-sensitive synthetics bound transfer. School-mean and multi-fold ablations, together with shared-head constant/shuffled/entropy gates, warn against over-attributing global wins to the entropy map alone. In practice, papers on interaction systems that mix recurrent motifs with exploratory edges should report entropy-conditioned errors rather than network-wide averages alone. In sum: an entropy-conditioned router can mitigate errors by suppressing message passing in selected strata, but the effect is neither universal across domains nor stable under all temporal generators. Multi-seed bootstrap intervals for contact Δ include zero; learned gates track the fixed gate on school contacts and improve on workplace contacts while logistic remains competitive in absolute log-loss; and shared-head constant mixers can approach entropy gating on some scenes. The claim is therefore scoped to a diagnostic evaluation protocol. Future work should tighten seed-controlled window contrasts, publish multi-seed unit-level intervals, add at least one full temporal-GNN backbone, and release anonymous code with time-index audits.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/e28111270/s1.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: SocioPatterns contact data are available from the SocioPatterns project under the stated non-commercial terms [4,6,31]. BenchTemp interaction streams (including MOOC, Contacts, and SocialEvo) are available via the BenchTemp release [32]. An anonymized repository with preprocessing scripts, fold boundaries, candidate sets, seeds, configuration files, table/figure builders, and a time-index unit test will be linked at submission; locked medium tables and comparison-suite outputs are included.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Holme, P.; Saramäki, J. Temporal Networks. *Phys. Rep.* **2012**, *519*, 97–125. [CrossRef]
2. Masuda, N.; Holme, P. Predicting and Controlling Infectious Disease Epidemics Using Temporal Networks. *F1000Prime Rep.* **2013**, *5*, 6. [CrossRef] [PubMed]

3. Cattuto, C.; Van den Broeck, W.; Barrat, A.; Colizza, V.; Pinton, J.F.; Vespignani, A. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE* **2010**, *5*, e11596. [CrossRef] [PubMed]
4. Fournet, J.; Barrat, A. Contact Patterns among High School Students. *PLoS ONE* **2014**, *9*, e107878. [CrossRef] [PubMed]
5. Stéhlé, J.; Voirin, N.; Barrat, A.; Cattuto, C.; Isella, L.; Pinton, J.F.; Quaggiotto, M.; Van den Broeck, W.; Régis, C.; Lina, B.; et al. High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *PLoS ONE* **2011**, *6*, e23176. [CrossRef] [PubMed]
6. Génois, M.; Vestergaard, C.L.; Fournet, J.; Panisson, A.; Bonmarin, I.; Barrat, A. Data on Face-to-Face Contacts in an Office Building Suggest a Low-Cost Vaccination Strategy Based on Community Linkers. *Netw. Sci.* **2015**, *3*, 326–347. [CrossRef]
7. Liben-Nowell, D.; Kleinberg, J. The Link-Prediction Problem for Social Networks. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1019–1031. [CrossRef]
8. Lü, L.; Zhou, T. Link Prediction in Complex Networks: A Survey. *Phys. A* **2011**, *390*, 1150–1170. [CrossRef]
9. Martínez, V.; Berzal, F.; Cubero, J.C. A Survey of Link Prediction in Complex Networks. *ACM Comput. Surv.* **2016**, *49*, 69. [CrossRef]
10. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*; ICLR: San Francisco, CA, USA, 2017.
11. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*; ICLR: San Francisco, CA, USA, 2018.
12. Rusch, T.K.; Bronstein, M.M.; Mishra, S. A Survey on Oversmoothing in Graph Neural Networks. *arXiv* **2023**, arXiv:2303.10993.
13. Kumar, S.; Zhang, X.; Leskovec, J. Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; ACM: New York, NY, USA, 2019; pp. 1269–1278. [CrossRef] [PubMed]
14. Rossi, E.; Chamberlain, B.; Frasca, F.; Eynard, D.; Monti, F.; Bronstein, M. Temporal Graph Networks for Deep Learning on Dynamic Graphs. In *Proceedings of the ICML Workshop on Graph Representation Learning*; PMLR: Cambridge, MA, USA, 2020.
15. Pareja, A.; Domeniconi, G.; Chen, J.; Ma, T.; Suzumura, T.; Kanezashi, H.; Kaler, T.; Schardl, T.; Leiserson, C. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Washington, DC, USA, 2020; Volume 34, pp. 5363–5370. [CrossRef]
16. Xu, D.; Ruan, C.; Korpeoglu, E.; Kumar, S.; Achan, K. Inductive Representation Learning on Temporal Graphs. In *Proceedings of the International Conference on Learning Representations (ICLR)*; ICLR: San Francisco, CA, USA, 2020.
17. Kazemi, S.M.; Goel, R.; Jain, K.; Kobyzev, I.; Sethi, A.; Forsyth, P.; Poupart, P. Representation Learning for Dynamic Graphs: A Survey. *J. Mach. Learn. Res.* **2020**, *21*, 2648–2720.
18. Li, A.; Pan, Y. Structural Information and Dynamical Complexity of Networks. *IEEE Trans. Inf. Theory* **2016**, *62*, 3290–3339. [CrossRef]
19. Sun, R.; Xin, Y.; Shao, Y.; Ju, P. Structural Entropy, Modal Diversity Entropy, and Accessibility Differentiation: A Study of the Western China–Central Asia Cross-Border Transportation Network. *Entropy* **2026**, *28*, 823. [CrossRef] [PubMed]
20. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [CrossRef] [PubMed]
21. Hawkes, A.G. Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika* **1971**, *58*, 83–90. [CrossRef]
22. An, L.; Dai, J. Transfer-Entropy- and Hawkes-Process-Driven Dynamic Measurement of Cross-Border Financial Risk Contagion in Directed, Weighted Networks. *Entropy* **2026**, *28*, 887. [CrossRef] [PubMed]
23. Tishby, N.; Pereira, F.C.; Bialek, W. The Information Bottleneck Method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*; IEEE: New York, NY, USA, 2000.
24. Wu, T.; Ren, H.; Li, P.; Leskovec, J. Graph Information Bottleneck. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 20437–20448.
25. Wu, Y.; Zhu, H. Conditional Information-Bottleneck Graph Clustering for Structured Representation Learning in Dynamic Vehicular ISAC Networks. *Entropy* **2026**, *28*, 884. [CrossRef] [PubMed]
26. Xi, Y.; Cui, X. Identifying Influential Nodes in Complex Networks Based on Information Entropy and Relationship Strength. *Entropy* **2023**, *25*, 754. [CrossRef] [PubMed]
27. Wang, B. Improving Multivariate Time-Series Anomaly Detection in Industrial Sensor Networks Using Entropy-Based Feature Aggregation. *Entropy* **2026**, *28*, 14. [CrossRef] [PubMed]
28. Kang, X.; Yu, T.; Wang, L.; Guo, Y.; Zhang, F. SS-AdaMoE: Spatio-Spectral Adaptive Mixture of Experts with Global Structural Priors for Graph Node Classification. *Entropy* **2026**, *28*, 355. [CrossRef] [PubMed]
29. Chen, Z. Graph Adaptive Attention Network with Cross-Entropy. *Entropy* **2024**, *26*, 576. [CrossRef] [PubMed]
30. Skarding, J.; Gabrys, B.; Musial, K. Foundations and Modeling of Dynamic Networks Using Dynamic Graph Neural Networks: A Survey. *IEEE Access* **2021**, *9*, 79143–79168. [CrossRef]
31. SocioPatterns Collaboration. SocioPatterns Datasets. 2012. Available online: https://sociopatterns.org/datasets (accessed on 15 November 2026).

32. Huang, Q.; Wang, X.; Rao, S.X.; Han, Z.; Zhang, Z.; He, Y.; Xu, Q.; Zhao, Y.; Zheng, Z.; Jiang, J. BenchTemp: A General Benchmark for Evaluating Temporal Graph Neural Networks. In *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering (ICDE)*; IEEE: New York, NY, USA, 2024; pp. 4044–4057. [CrossRef]
33. Rossetti, G.; Cazabet, R. Community Discovery in Dynamic Networks: A Survey. *ACM Comput. Surv.* **2018**, *51*, 35. [CrossRef]
34. Pastor-Satorras, R.; Vespignani, A. Epidemic Spreading in Scale-Free Networks. *Phys. Rev. Lett.* **2001**, *86*, 3200–3203. [CrossRef] [PubMed]
35. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
36. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2006.
37. Newman, M.E.J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]
38. Newman, M. *Networks*, 2nd ed.; Oxford University Press: Oxford, UK, 2018.
39. Guan, S.; Zhao, B.; Dong, Z.; Gao, M.; He, Z. GTAD: Graph and Temporal Neural Network for Multivariate Time Series Anomaly Detection. *Entropy* **2022**, *24*, 759. [CrossRef] [PubMed]
40. Zhao, H.; Wang, Y.; Duan, J.; Huang, C.; Cao, D.; Tong, Y.; Xu, B.; Bai, J.; Tong, J.; Zhang, Q. Multivariate Time-Series Anomaly Detection via Graph Attention Network. In *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM)*; IEEE: New York, NY, USA, 2020; pp. 841–850. [CrossRef]
41. Liu, X.; Sun, D.; Wei, W. Alleviating the Over-Smoothing of Graph Neural Computing by a Data Augmentation Strategy with Entropy Preservation. *Pattern Recognit.* **2022**, *132*, 108951. [CrossRef]
42. Watts, D.J.; Strogatz, S.H. Collective Dynamics of 'Small-World' Networks. *Nature* **1998**, *393*, 440–442. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.