Before It's Too Late: A State Space Model for the Early Prediction of Misinformation and Disinformation Engagement

Lin Tian
University of Technology Sydney
Sydney, Australia
lin.tian-3@uts.edu.au

Emily Booth
University of Technology Sydney
Sydney, Australia
emily.booth@uts.edu.au

Francesco Bailo
The University of Sydney
Sydney, Australia
francesco.bailo@sydney.edu.au

Julian Droogan
Macquarie University
Sydney, Australia
julian.droogan@mq.edu.au

Marian-Andrei Rizoiu
University of Technology Sydney
Sydney, Australia
marian-andrei.rizoiu@uts.edu.au

Abstract

In today's digital age, conspiracies and information campaigns can emerge rapidly and erode social and democratic cohesion. While recent deep learning approaches have made progress in modeling engagement through language and propagation models, they struggle with irregularly sampled data and early trajectory assessment. We present IC-Mamba🜂, a novel state space model that forecasts social media engagement by modeling interval-censored data with integrated temporal embeddings. Our model excels at predicting engagement patterns within the crucial first 15-30 minutes of posting (RMSE 0.118-0.143), enabling rapid assessment of content reach. By incorporating interval-censored modeling into the state space framework, IC-Mamba captures fine-grained temporal dynamics of engagement growth, achieving a 4.72% improvement over state-of-the-art across multiple engagement metrics (likes, shares, comments, and emojis). Our experiments demonstrate IC-Mamba's effectiveness in forecasting both post-level dynamics and broader narrative patterns (F1 0.508-0.751 for narrative-level predictions). The model maintains strong predictive performance across extended time horizons, successfully forecasting opinion-level engagement up to 28 days ahead using observation windows of 3-10 days. These capabilities enable earlier identification of potentially problematic content, providing crucial lead time for designing and implementing countermeasures. Code is available at: https://github.com/ltian678/ic-mamba. An interactive dashboard demonstrating our results is available at: https://ic-mamba.behavioral-ds.science/.

CCS Concepts

• **Information systems** → **Social networks**; • **Computing methodologies** → *Artificial intelligence*.

Keywords

State Space Model, Early Prediction, Interval-Censored, Information Propagation, Misinformation, Disinformation, Social Engagement.

1 Introduction

On 28 October 2017, an anonymous 4chan user made a brief yet impactful post on the platform claiming that Hillary Clinton was to be arrested in the coming days[1]. On 6 January 2021, devotees of then-outgoing President Donald Trump stormed the United States Capitol building in an act of domestic terrorism designed to prevent President-elect Joe Biden's election victory from being confirmed. Five people died during and in the immediate aftermath of the attack, and an additional four died in the subsequent months[2]; and over 140 police officers were injured[3]. Investigations by the Associated Press of the online social media profiles of over 120 of the rioters revealed high levels of adherence to the QAnon conspiracy theory that had begun just four years prior on 4chan[4]. This incident highlights how social media platforms can accelerate the spread of harmful content, particularly misinformation (false information shared without intent to harm) and disinformation (deliberately created and shared false information) [26, 36].

Given these ongoing impacts, the question must be asked: what if we could have seen this coming? More specifically, what if it had been possible to forecast user engagement with fringe ideologies before they morph into widespread movements? We introduce IC-Mamba🜂, a model capable of forecasting user engagement with online content. We go beyond the level of atomic posts to forecast the number of likes, shares, emoji reactions, and comments for "emerging opinions" – particular worldviews supported across a series of posts. Our framework can forecast the arrival rate of posts supporting an opinion, and forecast the engagement for each, obtaining estimates of the total level of engagement for the entire

---

[1]https://www.bellingcat.com/news/americas/2021/01/07/the-making-of-qanon-a-crowdsourced-conspiracy/
[2]https://www.factcheck.org/2021/11/how-many-died-as-a-result-of-capitol-riot/
[3]https://www.nytimes.com/2021/08/03/us/politics/capitol-riot-officers-honored.htm
[4]https://apnews.com/article/us-capitol-trump-supporters-1806ea8dc15a2c04f2a68acd6b55cace

Lin Tian, Emily Booth, Francesco Bailo, Julian Droogan, and Marian-Andrei Rizoiu
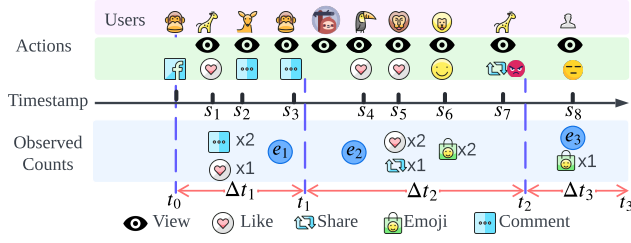
Figure 1: Illustration of interval-censored social media engagement data. Following a post's creation at t_0, users perform engagement actions (view, like, comment, share, emoji) at timestamps s_1 through s_8. While individual actions occur continuously, engagement data is only collected at discrete observation points t_j, where each engagement vector e_j captures the cumulative counts of different interaction types over intervals of length $\Delta t_j = t_{j+1} - t_j$.

Table 1: Information used in related studies and our work (IC-Mamba🌀). V: views, S: shares/retweets, C: comments, L: likes, E: emojis, WB: Weibo, YT: Youtube, FB: Facebook.

	V	S	C	L	E	platform
DeepCas [27]		✓				X
DeepHawkes [7]		✓				X
Topo-LSTM [41]		✓				X
HIP [35]	✓	✓				YT
DeepInf [32]		✓		✓		WB and X
B-Views [44]	✓					YT
SNPP [12]				✓		X
CasFlow [45]		✓				WB
MBPP [34]	✓	✓				YT
IC-TH [21]		✓				X
OMM [4]		✓				X,YT& FB
IC-Mamba🌀		✓	✓	✓	✓	FB

opinion. Our analysis leverages CrowdTangle [1] data with interval-censored engagement metrics, where observations are made at discrete time points with engagement counts recorded for each interval (as seen in Fig. 1).

Recent deep learning approaches have made progress in modeling social media engagement through different architectural innovations: language models to capture coordinated posting behaviors [2, 37–39] and propagation models to model information diffusion [19, 21–25, 30, 46, 47]. State space models have also demonstrated strong performance on sequential prediction tasks [11, 15], with their latent state representations theoretically well suited for temporal dependencies. However, these approaches face two key limitations when applied to mis/disinformation engagement forecasting: (1) they primarily focus on classification tasks rather than quantifying future temporal patterns of engagement, and (2) they struggle with the irregularly sampled nature of viral content.

Our main contributions address three research questions (RQs) at the intersection of temporal modeling and social media dynamics:

RQ1: *How can we effectively model irregular temporal patterns in social media engagement?*: Through IC-Mamba's time-aware embeddings and state space model architecture, we capture the dynamics of online interactions, achieving a 4.72% improvement over the state-of-the-art approaches.

RQ2:*Can we predict viral potential within the critical early window?* : IC-Mamba🌀 shows strong performance in the crucial 15-30 minute post-publication window (RMSE 0.118-0.143), while capturing both granular post-level dynamics and broader narrative patterns (F1 0.508-0.751 for narrative-level predictions).

RQ3: *How can we forecast engagement with emerging opinions early? Can we improve the accuracy and confidence of these forecasts as engagement data streams in over time?*: Our experiments show the model effectively forecasts engagement dynamics early, using 3-, 7-, and 10-day windows to predict spreading patterns up to 28 days, with performance improving as more data streams in.

As a tool, IC-Mamba🌀 streamlines the work of human experts, enabling earlier identification of problematic content, and therefore, providing more time to design and implement countermeasures.

2 Related Work

This section reviews relevant literature in two key areas that underpin our approach to modeling and predicting social engagements during outbreak events such as information operations and natural disasters like the infamous 2019-2020 Australian Bushfires: popularity and engagement prediction on social media platforms, and state space models for sequence modeling and prediction.

2.1 Popularity and Engagement Prediction

Social media engagement prediction research spans various platforms and prediction tasks. DeepCas [27] used random walk and attention mechanisms to predict final cascade size on X, while SNPP [12] applied temporal point process with a gated recurrent unit architecture for tweet repost count prediction. Topo-LSTM [41] incorporated user interaction sequences through a topological structure for retweet prediction, while DeepInf [32] on Weibo and X predicted user retweets, likes, and following behaviors using graph convolutional networks within the 2-hop neighborhood. CasFlow [45] used hierarchical attention networks for modeling Weibo reposts. Several approaches have focused on handling temporal dynamics in engagement prediction. DeepHawkes [7] integrated reinforcement learning with Hawkes processes for retweet cascade prediction. For YouTube, HIP [35] and MBPP [6, 34] advanced temporal modeling for views prediction, with B-Views [44] specifically addressing cold-start scenarios. Recently, IC-TH [21] tackled the challenge of incomplete observations in retweet prediction on X, while OMM [4] and BMH [5] proposed a mathematical framework for shares prediction on X, YouTube, and Facebook. While these approaches have advanced cascade modeling, existing interval-censored methods (IC-TH, MBPP) focus on single-post dynamics without considering broader opinion-level patterns. Our work extends beyond individual post predictions to model collective opinion engagement on Facebook, and we include these interval-censored models along with neural approaches (TH [49]) as baselines.

2.2 State Space Model in Sequence Modeling

State Space Models (SSMs) have recently emerged as a robust alternative to traditional sequence modeling approaches, particularly for long-range dependency capture [18, 33]. Subsequent work has showed their efficiency in processing extremely long sequences [10, 16] and competitive performance in language modeling [11, 15]. Despite these advancements, few studies have applied modern SSM architectures to predict or analyze social media engagement, particularly in the context of disinformation campaigns or crisis events. Most existing methods rely on graph-based [29], RNN [42], or transformer approaches [49] that typically assume uniform sampling or discret snapshots. Such assumptions often overlook fine-grained temporal patterns crucial to disinformation campaigns or crisis events. In contrast, modern SSMs can naturally handle non-uniform intervals and continuous-time dynamics, making them well-suited for rapidly unfolding social media processes. Our work bridges this gap by extending the Mamba architecture to handle non-uniform intervals, while identifying misinformation opinions and disinformation narratives.

3 Interval-Censored Mamba (IC-Mamba)

This section introduces IC-Mamba🌀, our proposed approach for engagement prediction illustrated in Fig. 2. We begin with the problem statement (Section 3.1) and then detail the key components of our architecture: the time-aware positional embeddings (Section 3.2), the content and sequence embeddings (Section 3.3), interval-censored state space modeling (Section 3.4), the pretraining strategies (Section 3.5), and the two-tier architecture that enables predictions at both post and opinion levels (Section 3.6).

3.1 Problem Statement

Let $\mathcal{E}$ denote a social outbreak event with associated posts $\mathcal{P} = \{p_1, p_2, \ldots, p_N\}$. For each post $p \in \mathcal{P}$, we define a tuple (t_0, x, u, o, H) where t_0 denotes the original posting time; x represents the textual content; u captures the user metadata; $o \in O$ indicates the opinion class from the set of possible opinions O; and the interval-censored engagement history is defined as $H = \{(t_j, e_j)\}_{j=1}^{m}$, with m as the total number of observation intervals. Each e_j is a d-dimensional vector capturing different types of engagement at observation time t_j, with intervals $\Delta t_j = t_{j+1} - t_j$ between consecutive observations – see also Fig. 1 for how these quantities interact. See Table 6 for a complete reference of mathematical notations used in this work.

Given an observation window τ_{obs} (e.g., 1 day), let $H_{\tau_{obs}}(p) = \{(t, e) \in H \mid t_0 \leq t \leq t_0 + \tau_{obs}\}$ denote the initial interval-censored engagement history. Let Δt be a fixed time interval (e.g., 5 minutes) and T be the prediction horizon (e.g., 28 days). Our goal is to predict the engagement trajectory at regular intervals: $\{\hat{e}(t_0 + \tau_{obs} + k\Delta t)\}_{k=1}^{K}$, where $K = \lfloor T/\Delta t \rfloor$ represents the number of prediction points.

Using this setup, we address two primary tasks. (1) Social Engagement Prediction: We predict engagement at both individual and collective levels. *Post level*: Predict the engagement trajectory $\hat{e}(t_0 + \tau_{obs} + k\,\tau_{step})_{k=1}^{K}$ at regular intervals τ_{step} up to horizon T (with $K = \lfloor T/\tau_{step} \rfloor$), as well as the total cumulative engagement over T. *Opinion level:* For a given opinion o, predict the collective

trajectory $\hat{E}_o(t_0 + \tau_{obs} + k\tau_{step})_{k=1}^{K}$, where $\hat{E}_o$ is the sum of engagements across all posts $\mathcal{P}_o$ expressing o. (2) Opinion Classification: We learn a mapping $f : (x, u, H_{\tau_{obs}}) \mapsto O$ that assigns a post to an opinion class based on its content x, user metadata u, and engagement history $H_{\tau_{obs}}$.

3.2 Time-aware Positional Embeddings

The temporal dynamics of social media engagement operate at multiple scales – from rapid initial spread to long-term influence patterns. To capture these multi-scale dynamics, we introduce a dual strategy featuring Relative Temporal Encoding (RTE) and Absolute Temporal Encoding (ATE). RTE captures temporal relationships between two time points t and t_{ref} as $RTE(t, t_{ref}) = \sin\left(\frac{t - t_{ref}}{\sigma}\right)$, where σ is a learnable parameter that allows the model to adapt to varying engagement velocities. ATE is capturing predictions to the global event timeline by mapping each time point t into a sinusoidal embedding space:

$$ATE(t) = \left[\sin\left(\frac{t}{10000^{2i/d}}\right), \cos\left(\frac{t}{10000^{2i/d}}\right)\right]_{i=0}^{d/2-1}.$$

These embeddings combine through a learnable projection:

$$PE(t, t_{ref}) = W_p \begin{bmatrix} RTE(t, t_{ref}) \\ ATE(t) \end{bmatrix},$$

which is then modulated by observed engagement $EPE(t, t_{ref}, e) = PE(t, t_{ref}) \odot (1 + \log(1 + e))$, where $\odot$ denotes element-wise multiplication, and e is the engagement vector at time t.

This engagement-sensitive embedding enables the model to learn characteristic temporal patterns associated with different levels of social impact. For each post $p \in \mathcal{P}$ and a prediction time τ_k, we construct a *time-aware embedding sequence* $TE^k(p) \in \mathbb{R}^{(m_k+1) \times d}$ as $TE^k(p) = [EPE(t_j, \tau_k, e_j) \mid (t_j, e_j) \in H_{\tau_{obs}}(p)] \cup [PE(\tau_k, \tau_k, 0)]$, where $H_{\tau_{obs}}(p) = \{(t, e) \in H \mid t_0 \leq t \leq t_0 + \tau_{obs}\}$ is the observed engagement history within the observation window τ_{obs}.

3.3 Content and Sequence Embedding

To create a unified representation of social media posts, we must handle both textual content and temporal patterns. We use a byte-level BPE tokenizer [3] to process the social media text, enabling us to embed the multi-modal information (content, user metadata, and temporal dynamics) into a single sequence representation: $SE(p) = Encoder([CLS] \oplus [x] \oplus [SEP] \oplus [u] \oplus [SEP] \oplus [T] \oplus [SEP] \oplus [e_j])$. Here, $Encoder$ is a transformer-based function, x is the post text, u is user metadata, $T = \{t_0, t_1, \ldots, t_m\}$ is the post's timeline of engagement events, $\{e_j\}$ are engagement counts, and $[CLS]$ and $[SEP]$ are special tokens. Note that the $Encoder$ function maps the input sequence to a fixed-dimensional space $\mathbb{R}^d$, where d is the embedding dimension. This allows for building uniform representations regardless of the posts' content or engagement history length.

3.4 Interval-Censored State Space Modeling

Here, we extend the Mamba architecture to incorporate time intervals within the state space model. Standard SSMs assume regular sampling intervals, which fails to capture social media engagement's irregular and censored nature (see Fig. 1). We address this
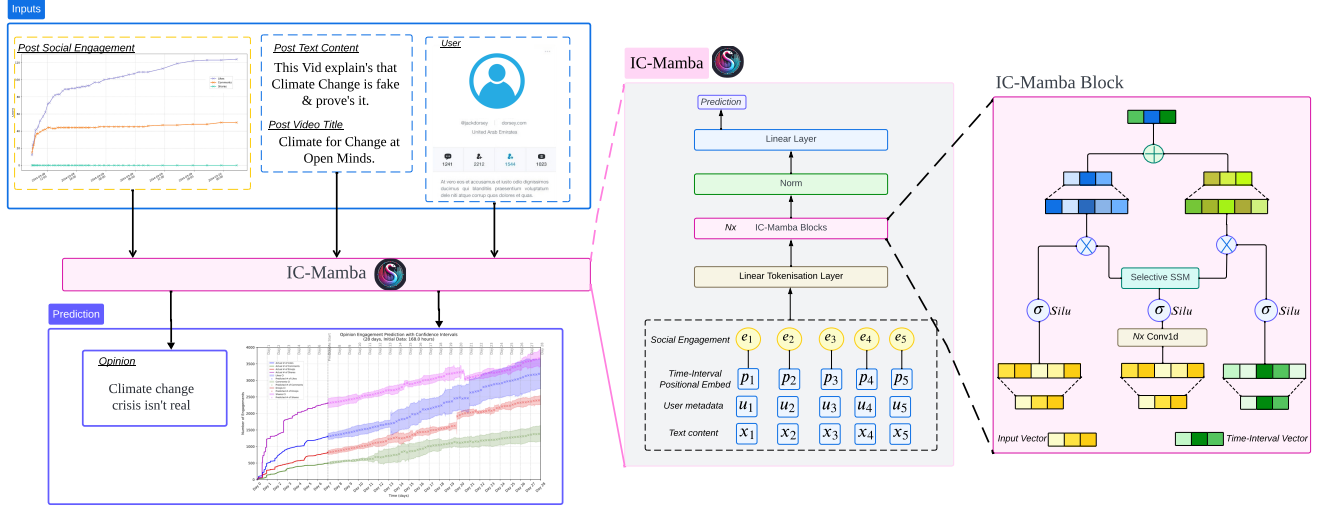
Figure 2: Overview of the IC-Mamba Architecture for social media engagement prediction. (left panel) The model first takes three types of inputs (interval-censored social engagement, post content, and user metadata). These inputs are tokenized through a linear tokenization layer. The tokenized sequence (combination of temporal embedding, positional embeddings and user embeddings) is processed through N-stacked IC-Mamba⑤ blocks. (right panel) Each IC-Mamba⑤ block contains a selective SSM mechanism and parallel Conv1d operations to handle input and time-interval vectors simultaneously. Lastly, the processed features go through normalization and linear layers to generate the final social engagement predictions.

through three key components: interval-aware state representation, time-dependent transitions, and selective state updates.

Interval-aware State Representation. For each observation time t_j in the engagement history $H_{\tau_{obs}}(p)$, we construct an interval-aware vector $v_j \in \mathbb{R}^{4d}$:

$$v_j = [\Delta t_j^-; \log(1 + e_j); \Delta t_j^+; \log(1 + \hat{e}_{j+1})],$$

where $\Delta t_j^- = t_j - t_{j-1}$ captures the time since the last observation, e_j is the current engagement vector, $\Delta t_j^+ = t_{j+1} - t_j$ is the forward interval length, and $\hat{e}_{j+1}$ is the predicted next engagement vector.

To maintain a consistent representation when transitioning from variable-length historical intervals to fixed-length prediction intervals, at each prediction time point τ_k, we construct: $v_k = [\tau_k - t_j; \log(1 + e_j); \tau_{k+1} - \tau_k; \log(1 + \hat{e}_k)]$, using the last observed engagement (t_j, e_j) in $H_{\tau_{obs}}$.

Time-Dependent State Transitions. We handle varying-length censored intervals by modifying the standard SSM architecture to incorporate time-dependent state transitions. For a hidden state dimension D_h and input dimension D, our model becomes:

$$\mathbf{A}_t(\Delta t) = \exp(\Delta t \cdot \tilde{\mathbf{A}}_t) \in \mathbb{R}^{D_h \times D_h},$$

$$\mathbf{h}_t = \mathbf{A}_t(\Delta t)\mathbf{h}_{t-1} + \mathbf{B}_t \mathbf{x}_t, \qquad \mathbf{y}_t = \mathbf{C}_t^T \mathbf{h}_t,$$

where $\mathbf{h}_t \in \mathbb{R}^{D_h}$ is the hidden state at time t, $\mathbf{x}_t \in \mathbb{R}^D$ is derived from the interval-aware vector v_j, and the matrix exponential $\exp(\Delta t \cdot \tilde{\mathbf{A}}_t)$ enables smooth interpolation across censored intervals.

Selective State Processing. We integrate the temporal embeddings $(TE^k(p))$ and interval-aware vectors through parallel pathways:

$$[\mathbf{X}, \Delta, \mathbf{B}, \mathbf{C}] = \text{Projection}\left(\mathbf{V}, TE^k(p)\right) \in \mathbb{R}^{L \times (D+1+2N)}$$

where L is the sequence length, $\mathbf{V} \in \mathbb{R}^{L \times 4d}$ is the sequence of interval-aware vectors, and $TE^k(p)$ provides temporal context. The selective SSM mechanism then processes as follows:

$$\mathbf{Y} = \text{SSM}(\tilde{\mathbf{A}}, \mathbf{B}, \mathbf{C}, \mathbf{X}, \Delta) \in \mathbb{R}^{L \times D},$$

The final output is modulated through a gating mechanism:

$$\text{Output} = \mathbf{Y} \odot \sigma\left(\text{Conv1d}(\mathbf{X})\right) \in \mathbb{R}^{L \times D},$$

where σ is the Silu activation function [14] and Conv1d [17] captures local engagement patterns.

3.5 IC-Mamba Pretraining

Creating labeled sets of misinformation and disinformation campaigns is a human-time-intensive process, and often, the resulting training sets are too small to allow training an architecture such as IC-Mamba⑤ from scratch. Algorithm 1 outlines the pretraining procedure for IC-Mamba⑤. We introduce $\mathcal{D} = \{(p_i, H_i, x_i, u_i)\}_{i=1}^M$, a pretraining dataset comprising 1.78 million posts and their associated social engagement timelines – totaling over 153 million timelines – collected from the two datasets SocialSense [20] and DiN (detailed in Section 4.1). Here M is the number of posts and $H_i = \{(t_{i,n}, e_{i,n})\}_{n=1}^{m_i}$ with $|H_i| = m_i$ represents the complete engagement history for post p_i.

Objective Function. We define two objective functions that we combine for pretraining.

Engagement Prediction Loss. For each post, we train the model to predict the next engagement vector:

$$\mathcal{L}_{\text{pred}} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \sum_{j=0}^{m-1} \|\hat{e}_{j+1} - e_{j+1}\|^2 , \tag{1}$$

Algorithm 1 IC-Mamba🌀 Pretraining

1: Initialize parameters $\theta = \{\tilde{\mathbf{A}}, \mathbf{B}, \mathbf{C}, \mathbf{W}_p, \theta_{\text{Encoder}}\}$
2: **for** epoch = 1 to N_{epochs} **do**
3:      **for** batch $\mathcal{B} \subset \mathcal{D}$ **do**
4:          Construct interval-aware vectors $\{\mathbf{v}_j\}_{j \in \mathcal{B}}$
5:          Compute temporal embeddings $\{TE^k(p)\}_{p \in \mathcal{B}}$
6:          $[\mathbf{X}, \mathbf{\Delta}, \mathbf{B}, \mathbf{C}] \leftarrow \text{Projection}(\{\mathbf{v}_j\}, \{TE^k(p)\})$
7:          $\mathbf{H} \leftarrow \text{SSM}(\tilde{\mathbf{A}}, \mathbf{B}, \mathbf{C}, \mathbf{X}, \mathbf{\Delta})$
8:          $\hat{\mathbf{E}} \leftarrow \text{MLP}(\mathbf{H})$
9:          Compute $\mathcal{L}_{\text{pred}}$ (Eq. (1)) and $\mathcal{L}_{\text{temp}}$ (Eq. (2))
10:         Update θ using $\nabla_\theta(\mathcal{L}_{\text{pred}} + \lambda \mathcal{L}_{\text{temp}})$
11:     **end for**
12: **end for**
13: **return** θ

where $\hat{e}_{j+1} \in \mathbb{R}^d$ is the predicted engagement vector.

Temporal Coherence Loss. We enforce consistent state transitions across intervals:

$$\mathcal{L}_{\text{temp}} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \sum_{j=0}^{m-1} \|\mathbf{h}_{j+1} - \exp(\Delta t_j^+ \cdot \tilde{\mathbf{A}}_t)\mathbf{h}_j\|^2 , \qquad (2)$$

where $\mathbf{h}_j \in \mathbb{R}^{D_h}$ is the hidden state at time t_j and the exponential term comes from our SSM formulation.

The pretraining loss combines these objectives from Eq. (1) and Eq. (2) as $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pred}} + \lambda \mathcal{L}_{\text{temp}}$, where λ is a hyperparameter balancing the two losses.

3.6 Two-Tier IC-Mamba🌀 Architecture

It is desirable to model and predict the engagement dynamics of a group of posts expressing the same opinion – dubbed *the engagement of an opinion*. We propose a hierarchical two-tier architecture, showcased in Fig. 3. The intuition of the two-tier IC-Mamba🌀 model is that the first tier (IC-Mamba$_1$) models the arrival of engagement on an individual post. The second tier (IC-Mamba$_2$) models the arrival of posts within an opinion.

Post-Level Processing. In the first tier, for each opinion o, we process all posts $p_i \in \mathcal{P}_o$ individually using the IC-Mamba1 model:

$$\mathbf{h}_i = \text{IC-Mamba}_1(H_{\tau_{obs}}(p_i), x_i, u_i), \quad \forall p_i \in \mathcal{P}_o \qquad (3)$$

where $H_{\tau_{obs}}(p_i)$ is the interval-censored engagement over observation window and $\mathbf{h}_i$ the hidden state representation of post p_i.

Group-Level Dynamics. In the second tier, we model the temporal interactions between posts sharing opinion o. By ordering posts in $\mathcal{P}_o$ chronologically by posting time t_i^p, we capture the inter-post intervals $\delta t_i = t_{i+1}^p - t_i^p$ between posts in the group. The group-level dynamics are modeled using IC-Mamba$_2$ with $\mathbf{h}_i$ from Eq. (3):

$$z_o = \text{IC-Mamba}_2((\mathbf{h}_i, \delta t_i)).$$

4 Experiments and Results

In this section, we present the experimental setup and the results we obtain; including datasets and data insights (Section 4.1), the baseline models we compare against (Section 4.2), and the results that address our research questions (Section 4.3).
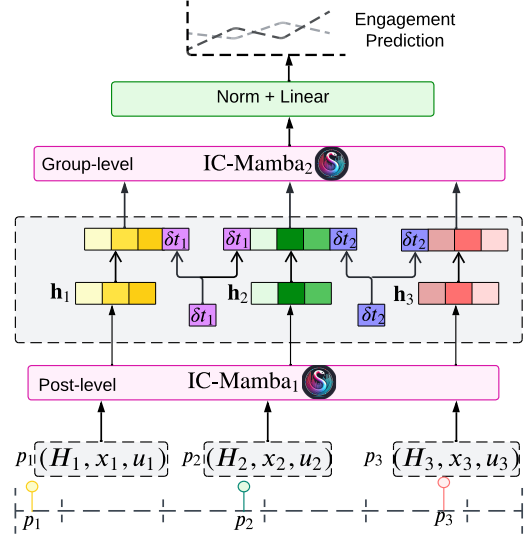
Figure 3: Two-Tier IC-Mamba🌀 Architecture. The bottom-tier model (IC-Mamba$_1$) learns post-level representations from historical (H), content (x), and user (u) features, while the top-tier model (IC-Mamba$_2$) captures temporal dependencies across intervals δt to jointly predict individual post virality and aggregate narrative engagement dynamics.

Table 2: Dataset Statistics

| Dataset	Bushfire	Climate Ch.	Vaccin.	COVID-19	DiN
#posts	78,030	138,278	178,894	640,100	746,653
#users	13,438	25,850	34,652	67,727	41
#opinions	15	24	27	17	9

4.1 Datasets

Datasets. Our experiments use two Facebook datasets: the theme-focused SocialSense dataset [20] and the user-centric Disinformation Network (DiN) dataset. For each post in our datasets, we collect historical engagement metrics (likes, shares, comments, emoji reactions) collected via CrowdTangle API[5]. *SocialSense* contains posts and comments from four main themes during 2019-2021(see Table 2) that attracted significant volumes of misinformation and conspiratorial discussions. The *DiN dataset* comprises posts from 41 accounts (2019-2024). Social science experts systematically analyzed and assigned narrative labels to these posts through comprehensive content evaluation to detect suspected coordinated information operations. The two datasets capture the dynamics of misinformation across diverse real-world events (SocialSense) and disinformation narrative spread by information operation networks (DiN). [6]

Data Insights. Fig. 4(a) and (b) present the Empirical Complementary Cumulative Distribution Functions (ECCDFs) for likes,

---

[5]https://www.crowdtangle.com/ before its termination in August 2024.
[6]Note that, posts with fewer than four engagement intervals were excluded from model evaluation to ensure sufficient temporal depth.
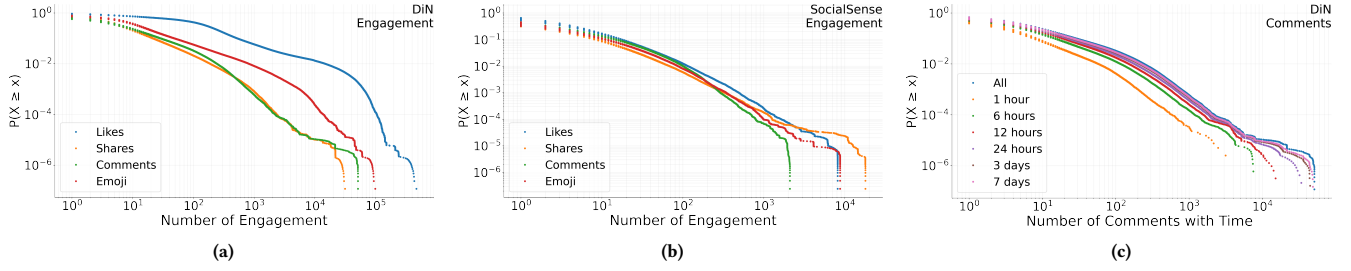
Figure 4: Engagement distribution patterns across social media content. (a) Log-scale ECCDF of engagement metrics for the DiN dataset. (b) Log-scale ECCDF of engagement metrics from the climate change theme in SocialSense. (c) Temporal evolution of comment distributions across different time windows ranging from 1 hour to 7 days. Note: ECCDF represents Empirical Complementary Cumulative Distribution Functions.

shares, comments, and emoji reactions across DiN (a) and the Climate Change theme in SocialSense (b). The survival probability $P(X \geq k)$ measures the likelihood of achieving at least k engagements [9], and the power-law exponent α characterizes the decay rate [31]. While Climate Change content rarely exceeds 10^4 total engagements, DiN reaches 10^6, indicating significantly broader reach. In the low-engagement regime ($1 \leq k \leq 10$), DiN exhibits a higher survival probability ($\alpha \approx 2.1$) compared to Climate Change ($\alpha \approx 2.4$), suggesting stronger early visibility potential. The mid-range ($10 \leq k \leq 1000$) shows uniform decay across engagement types for Climate Change, reflecting organic interaction patterns. In contrast, DiN reveals marked stratification, especially in likes. Beyond $k > 1000$, Climate Change content plateaus near 10^3 engagements, aligning with established social network theory regarding human-scale constraints – approximately 150 stable connections, known as Dunbar's number [13] – while DiN content transcends these natural limits, reaching 10^6 engagements.

Fig. 4(c) offers examines comment distributions over time windows ranging from one hour to seven days. The scale-invariant, power-law structure persists across all observation periods, though longer windows (3–7 days) exhibit slightly elevated survival probabilities beyond 10^3. This self-similar temporal behavior distinguishes naturally diffusing, high-visibility content from artificially amplified patterns, underscoring the unique viral longevity of DiN.

4.2 Baselines and Experimental Setup

We compare our IC-Mamba🝢 model against the following state-of-the-art baselines, including generative models, transformer-based architectures and state space models:

- **TimeSeriesTransformer** [28][7] is a transformer-based model specifically adapted for time series forecasting. It applies self-attention mechanisms to capture temporal dependencies.
- **Informer** [48][8] is a long sequence time-series forecasting model that uses a ProbSparse self-attention mechanism to handle long-term dependencies.
- **Autoformer** [43][9] is a decomposition-based architecture for long-term time series forecasting. It uses an auto-correlation

mechanism to identify period-based dependencies and a series decomposition architecture for trend-seasonal decomposition.

- **Mean Behaviour Poisson (MBP)** [34] is a generative time series model that uses a compensator function to model non-linear engagement patterns. It treats each engagement as an event in a continuous time process and optimizes post-specific parameters to model the expected cumulative engagements over time to capture the growth patterns.
- **Transformer-Hawkes (TH)** [49] is a model that combines the transformer architecture with the Hawkes process for modeling sequential events. It uses self-attention mechanisms to capture temporal dependencies in event sequences.
- **Interval-Censored Transformer Hawkes (IC-TH)** [21] is a TH extension designed to handle interval-censored data. It adapts the transformer architecture to work with event data where exact occurrence times are unknown but bounded within intervals.
- **TS-Mixer** [8][10] is a model that combines MLPs and transformers for time series forecasting. It uses separate mixing operations across the temporal and feature dimensions, allowing it to capture both temporal patterns and feature interactions.
- **Mamba** [11][11] is a selective state space model, it uses selective algorithms instead of attention mechanisms for sequence modeling. It can handle long-range dependencies in sequential data for time series analysis tasks.

Experimental settings. We use a temporal holdout evaluation protocol across all the datasets. We chronologically order all posts and use the earliest 70% for training, the next 15% for validation, and the most recent 15% for testing. This ensures no future information leaks into training and models are evaluated on their ability to generalize to future posts. Models are implemented using PyTorch, with hyperparameters and other settings detailed in Appendix B.

4.3 Engagement Prediction–RQ1

We evaluate the performance of our models with two tasks: *engagement forecasting* and *opinion classification*. For *engagement prediction*, we observe the first six hours of engagement metrics for each post and forecast the overall engagement metrics (i.e., at

---

[7]https://huggingface.co/docs/transformers/en/model_doc/time_series_transformer
[8]https://huggingface.co/docs/transformers/en/model_doc/informer
[9]https://huggingface.co/docs/transformers/en/model_doc/autoformer

[10]https://github.com/google-research/google-research/tree/master/tsmixer
[11]https://huggingface.co/docs/transformers/en/model_doc/mamba2

Table 3: Post-level engagement prediction performance of IC-Mamba⑤ vs baselines on SocialSense (four themes) and DiN; measured using RMSE and MAPE (lower is better), and R^2 (higher is better). Best performance in boldface.

Model	Bushfire			Climate Change			Vaccination			COVID-19			DiN		
	RMSE	MAPE	R^2	RMSE	MAPE	R^2	RMSE	MAPE	R^2	RMSE	MAPE	R^2	RMSE	MAPE	R^2
TSTransformer [40]	0.185	0.232	0.651	0.192	0.241	0.643	0.180	0.226	0.658	0.188	0.236	0.647	0.221	0.276	0.568
Informer [48]	0.172	0.216	0.678	0.179	0.224	0.670	0.167	0.210	0.685	0.175	0.220	0.674	0.206	0.258	0.598
Autoformer [43]	0.163	0.204	0.697	0.169	0.212	0.689	0.158	0.198	0.704	0.166	0.208	0.693	0.195	0.244	0.619
MBPP [34]	0.183	0.229	0.655	0.192	0.241	0.643	0.181	0.227	0.656	0.189	0.237	0.645	0.222	0.278	0.566
IC-TH [21]	0.156	0.195	0.712	0.162	0.203	0.704	0.151	0.189	0.719	0.159	0.199	0.708	0.187	0.234	0.636
TH [49]	0.149	0.187	0.726	0.155	0.194	0.718	0.144	0.181	0.733	0.152	0.190	0.722	0.179	0.224	0.652
TS-Mixer [8]	0.155	0.194	0.714	0.161	0.202	0.706	0.150	0.188	0.721	0.158	0.198	0.710	0.186	0.233	0.638
Mamba [11]	0.124	0.155	0.776	0.129	0.161	0.768	0.119	0.149	0.783	0.127	0.159	0.772	0.150	0.188	0.708
IC-Mamba⑤ w/o text	0.123	0.154	0.778	0.128	0.160	0.770	0.118	0.148	0.785	0.126	0.158	0.774	0.145	0.181	0.718
IC-Mamba⑤ w/o user	0.120	0.150	0.785	0.125	0.156	0.777	0.115	0.144	0.792	0.123	0.154	0.781	0.146	0.183	0.716
IC-Mamba⑤ w/o time	0.121	0.151	0.783	0.126	0.157	0.775	0.116	0.145	0.790	0.124	0.155	0.779	0.149	0.186	0.711
IC-Mamba⑤	**0.118**	**0.148**	**0.789**	**0.123**	**0.154**	**0.781**	**0.113**	**0.142**	**0.796**	**0.121**	**0.152**	**0.785**	**0.143**	**0.179**	**0.723**

$T = \infty$). For *opinion classification*, we evaluate our model's classification performance at multiple granularities. We perform a *post-level opinion classification* across the four SocialSense themes (*bushfire, climate change, vaccination,* and *COVID-19*) – that is, we predict if a given post expresses one of the predefined opinions. For the DiN dataset, we perform a *user-level opinion classification* – classify the presence of opinions across multiple posts from the same user.

Post-level Engagement Prediction Performance – RQ1 Table 3 reports the performance metrics using three standard measures. We evaluate the models using RMSE to assess absolute prediction errors (crucial for high-engagement posts), MAPE for scale-independent accuracy, and R^2 to measure explained variance in engagement predictions. IC-Mamba⑤ outperforms all baselines on every metric (RMSE, MAPE, and R^2) and dataset, while the original Mamba architecture ranks consistently second, confirming the effectiveness of state space models. Among transformers, IC-TH improves upon TH, and TS-Mixer outperforms both Autoformer and Informer; TSTransformer lags behind. Interestingly, the lightweight MBP model, still competes well on some events (particularly *bushfire* and *climate change*). All models exhibit performance degradation on the DiN dataset, reflecting the complexity of predicting engagement in coordinated campaigns. For models supporting dynamic prediction time points, additional results for next-time and next social engagement metrics are provided in Appendix D.

We conduct an ablation study to understand the contribution of different components by removing text, user, and temporal features from IC-Mamba⑤ (Table 3). Text features demonstrate a stronger influence on SocialSense datasets, where their removal leads to a 0.005 RMSE increase, compared to a smaller 0.002 RMSE increase in the DiN dataset. This difference highlights the crucial role of textual content in organic content spread versus coordinated campaigns. Temporal features, conversely, show greater impact on the DiN dataset, where their removal results in a 0.006 RMSE increase, compared to a 0.003 RMSE increase in SocialSense datasets. This may suggest the strategic temporal patterns in coordinated disinformation campaigns. User features maintain consistent importance

across both datasets, with their removal causing similar performance degradation (0.002-0.003 RMSE increase) regardless of the dataset type. Even with text features removed, IC-Mamba⑤ still outperforms IC-TH, improving RMSE from 0.156 to 0.123 on the Bushfire dataset, demonstrating the fundamental strength of our model's architectural design.

Opinion-level Classification Performance –RQ1 In our classification settings, we tackled datasets of varying complexity: the *bushfire* dataset contains 9 opinions, *climate change* and *vaccination* each have 12 classes, the *COVID-19* dataset includes 10 classes, and the DiN (Disinformation Narrative) dataset comprises 9 distinct narrative labels. We also include a random classification baseline with an expected F1 score of $1/N$ for each dataset, where N is the number of classes. Note that we removed opinions with less than 5,000 posts in this experimental setting.

Table 4 presents the macro-averaged F1 scores for classification across models and datasets. IC-Mamba consistently outperforms all others, achieving F1 scores between 0.69 and 0.75. While BERT performs well on SocialSense (F1: 0.62–0.68), both models see significant drops on DiN, with IC-Mamba scoring 0.52 and BERT falling to 0.11. This highlights the limitations of text-only analysis for DiN, where narrative elements demand more complex temporal or contextual understanding. Informer, Autoformer, and Mamba struggle on SocialSense (F1 < 0.41) but perform relatively better on DiN, with Mamba achieving its best score of 0.32. This suggests that temporal and non-textual features are critical for narrative detection, contrasting with the outbreak event focus of SocialSense.

4.4 Early Engagement Prediction–RQ2

We vary the length of the observed period in the temporal holdout setup (see Section 4.2) to assess how well different models can forecast engagement in the critical initial hours after a post is made. Fig. 5a shows RMSE-based early prediction performance for the climate change theme in SocialSense, measured at intervals from

(a)   (b)   (c)
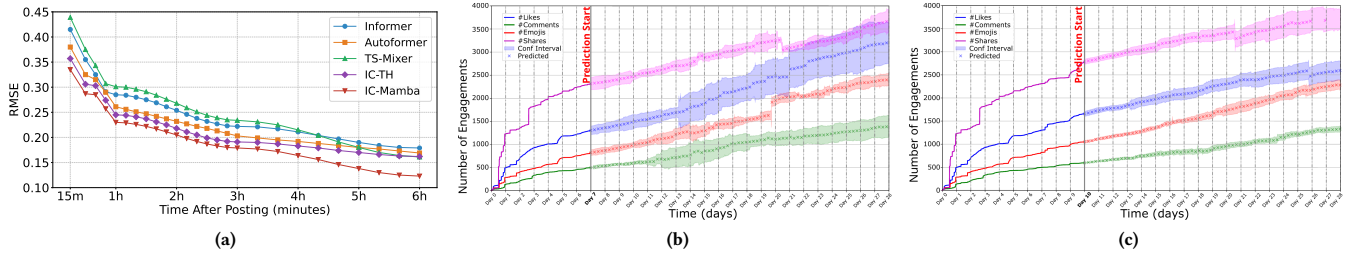
Figure 5: Comparative analysis of early prediction performance and dynamic forecasting. (a) Performance comparison on RMSE between IC-Mamba and baseline models from 15 minutes to 6 hours after posting. (b)(c) IC-Mamba🜨's 28-day predictions with 5-minute intervals using 7-day (b) and 10-day (c) input windows respectively.

Table 4: Opinion Classification results; F1 scores are reported; higher is better; best results in boldface.

Model	Bushfire	Climate	Vaccination	COVID-19	DiN
Random	0.111	0.083	0.083	0.083	0.111
Informer [48]	0.323	0.291	0.274	0.299	0.248
Autoformer [43]	0.342	0.313	0.278	0.324	0.255
BERT [40]	0.676	0.652	0.621	0.644	0.107
Mamba [11]	0.412	0.375	0.363	0.388	0.316
IC-Mamba🜨	**0.751**	**0.724**	**0.687**	**0.705**	**0.508**

15 minutes to 6 hours after posting, across Informer, Autoformer, TS-Mixer, IC-TH, and IC-Mamba🜨.

All models demonstrate substantial improvement in prediction accuracy over time, with error rates decreasing from 15 minutes to 6 hours. The most notable improvements occur in the first hour, particularly between 15-50 minutes, suggesting that the first hour of a post's life is crucial for accurate engagement forecasting. IC-Mamba🜨 outperforms other models across all time points, and its performance advantage increases over time. While all models show similar patterns of improvement in the first hour, IC-Mamba🜨 continues to achieve increasingly better RMSE scores through the 6-hour mark, reaching the lowest RMSE of 0.118. IC-TH maintains second-best performance throughout most of the timeline, followed by Autoformer. The Informer and TS-Mixer models show higher error rates, with their performance plateauing more quickly than the interval-censored approaches. This performance gap may illustrate the benefits of interval-censored modeling in engagement prediction tasks on real-world social media platforms, while the widening gap in RMSE scores over time suggests that IC-Mamba🜨's improvements go beyond just interval-censored modeling, potentially indicating better long-range dependency learning.

4.5 Dynamic Opinion-level Prediction–RQ3

This section simulates a real-world monitoring and forecasting scenario. We analyze the opinion "Climate change is a UN hoax" from the SocialSense climate dataset. Fig. 5(b)(c) demonstrates our dynamic prediction approach at opinion-level across multiple interaction types (likes, comments, emojis, and shares) over a 28-day

period. We showcase two scenarios of initial data windows – 168 hours (1 week), and 240 hours (10 days). Our model first processes the initial historical data window to establish baseline engagement patterns. As time progresses beyond these initial periods (marked by "Predictions Start" lines), the model continuously incorporates new engagement data to refine its predictions. The shaded areas around each prediction line represent the 95% confidence intervals – obtained from all previous prediction for this time – providing a measure of prediction uncertainty over time. We see that the uncertainty reduces as more initial data is available, suggesting that increased historical data improves the model's predictive accuracy.

5 Conclusion

IC-Mamba🜨 demonstrates strong performance in modeling interval-censored engagement data, providing early predictions of viral content, and tracking long-term opinion spread across platforms. Through the novel integration of interval-censored modeling and temporal embeddings within a state space model, IC-Mamba🜨 achieves strong performance in predicting dynamic misinformation and disinformation engagement patterns and opinion classification. These capabilities enable platforms and researchers to identify potentially harmful content and coordinated campaigns in their early stages, facilitating proactive intervention strategies while respecting platform constraints and user privacy. Future work could enhance IC-Mamba through cross-platform dynamics modeling, interpretable attention mechanisms, and real-time deployment adaptations. Online misinformation and information campaigns are part of our digital ecosystem today, but we do not need to resign ourselves to reactively attempting damage control after the fact. With IC-Mamba, we can identify the next QAnon or climate change denialism conspiracies before they gain mass exposure; and we could mitigate the damage it can do to our lives and democratic societies. We provide detailed discussion of ethical considerations and safeguards in Appendix A.

References

[1] [n. d.]. CrowdTangle. https://www.crowdtangle.com/
[2] Atanas Atanasov, Gianmarco De Francisci Morales, and Preslav Nakov. 2019. Predicting the Role of Political Trolls in Social Media. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*. 1023–1034.
[3] Sidney Black, Stella Biderman, Eric Hallahan, Quentin Gregory Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, Michael Martin Pieler, USVSN Sai Prashanth, Shivanshu Purohit, Laria Reynolds, Jonathan Tow, Ben Wang, and Samuel Weinbach. 2022. GPT-NeoX-20B: An Open-Source Autoregressive Language Model. In *Challenges & Perspectives in Creating Large Language Models*.
[4] Pio Calderon, Rohit Ram, and Marian-Andrei Rizoiu. 2024. Opinion Market Model: Stemming Far-Right Opinion Spread Using Positive Interventions. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 177–190.
[5] Pio Calderon and Marian-Andrei Rizoiu. 2024. *What Drives Online Popularity: Author, Content or Sharers? Estimating Spread Dynamics with Bayesian Mixture Hawkes*. 142–160. https://doi.org/10.1007/978-3-031-70362-1_9
[6] Pio Calderon, Alexander Soen, and Marian-Andrei Rizoiu. 2025. Linking Across Data Granularity: Fitting Multivariate Hawkes Processes to Partially Interval-Censored Data. *IEEE Transactions on Computational Social Systems* 12 (2 2025), 25–37. Issue 1. https://doi.org/10.1109/TCSS.2024.3486117
[7] Qi Cao, Huawei Shen, Keting Cen, Wentao Ouyang, and Xueqi Cheng. 2017. Deephawkes: Bridging the Gap between Prediction and Understanding of Information Cascades. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1149–1158.
[8] Si-An Chen, Chun-Liang Li, Sercan O Arik, Nathanael Christian Yoder, and Tomas Pfister. 2023. TSMixer: An All-MLP Architecture for Time Series Forecast-ing. *Transactions on Machine Learning Research* (2023).
[9] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. 2009. Power-law distributions in empirical data. *SIAM Review* 51, 4 (2009), 661–703.
[10] Tri Dao, Dan Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. 2022. Flashattention: Fast and Memory-Efficient Exact Attention with IO-Awareness. *Advances in Neural Information Processing Systems* 35 (2022), 16344–16359.
[11] Tri Dao and Albert Gu. 2024. Transformers are SSMs: Generalized Models and Efficient Algorithms Through Structured State Space Duality. In *Proceedings of the 41st International Conference on Machine Learning*, Vol. 235. 10041–10071.
[12] Keyan Ding, Ronggang Wang, and Shiqi Wang. 2019. Social Media Popularity Prediction: A Multiple Feature Fusion Approach with Deep Neural Networks. In *Proceedings of the 27th ACM International Conference on Multimedia*. 2682–2686.
[13] Robin IM Dunbar. 1992. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 22, 6 (1992), 469–493.
[14] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. 2018. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks* 107 (2018), 3–11.
[15] Albert Gu and Tri Dao. 2024. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In *Conference on Language Modeling*.
[16] Albert Gu, Karan Goel, and Christopher Re. 2022. Efficiently Modeling Long Sequences with Structured State Spaces. In *International Conference on Learning Representations*.
[17] Albert Gu, Isys Johnson, Karan Goel, Khaled Saab, Tri Dao, Atri Rudra, and Christopher Ré. 2021. Combining Recurrent, Convolutional, and Continuous-time Models with Linear State Space Layers. *Advances in Neural Information Processing Systems* 34 (2021), 572–585.
[18] Ramin Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. 2021. Liquid Time-Constant Networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 7657–7666.
[19] Jane Im, Eshwar Chandrasekharan, Jackson Sargent, Paige Lighthammer, Taylor Denby, Ankit Bhargava, Libby Hemphill, David Jurgens, and Eric Gilbert. 2020. Still out there: Modeling and Identifying Russian Troll Accounts on Twitter. In *12th ACM Conference on Web Science*. 1–10.
[20] Quyu Kong, Emily Booth, Francesco Bailo, Amelia Johns, and Marian-Andrei Rizoiu. 2022. Slipping to the Extreme: A Mixed Method to Explain How Extreme Opinions Infiltrate Online Discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 524–535.
[21] Quyu Kong, Pio Calderon, Rohit Ram, Olga Boichak, and Marian-Andrei Rizoiu. 2023. Interval-censored Transformer Hawkes: Detecting Information Operations using the Reaction of Social Systems. In *Proceedings of the ACM Web Conference 2023*. 1813–1821.
[22] Quyu Kong, Rohit Ram, and Marian-Andrei Rizoiu. 2021. Evently: A Toolkit for Analyzing Online Users via Reshare Cascade Modeling. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (New York, NY, USA). ACM, 1097–1100. https://doi.org/10.1145/3437963.3441708
[23] Quyu Kong, Marian-Andrei Rizoiu, Siqi Wu, and Lexing Xie. 2018. Will This Video Go Viral? Explaining and Predicting the Popularity of Youtube Videos. In *Companion of the The Web Conference 2018 on The Web Conference 2018 (WWW '18)* (Lyon, France). ACM Press, 175–178. https://doi.org/10.1145/3184558.3186972
[24] Quyu Kong, Marian-Andrei Rizoiu, and Lexing Xie. 2020. Describing and Predicting Online Items with Reshare Cascades via Dual Mixture Self-exciting Processes. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (New York, NY, USA). ACM, 645–654. https://doi.org/10.1145/3340531.3411861
[25] Quyu Kong, Marian-Andrei Rizoiu, and Lexing Xie. 2020. Modeling Information Cascades with Self-exciting Processes via Generalized Epidemic Models. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (New York, NY, USA). ACM, 286–294. https://doi.org/10.1145/3336191.3371821
[26] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.
[27] Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. 2017. Deepcas: An End-to-end Predictor of Information Cascades. In *Proceedings of the 26th International Conference on World Wide Web*. 577–586.
[28] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. 2019. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. *Advances in Neural Information Processing Systems* 32 (2019).
[29] Xiaodong Lu, Shuo Ji, Le Yu, Leilei Sun, Bowen Du, and Tongyu Zhu. 2023. Continuous-time graph learning for cascade popularity prediction. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. 2224–2232.
[30] Luca Luceri, Valeria Pantè, Keith Burghardt, and Emilio Ferrara. 2024. Unmasking the Web of Deceit: Uncovering Coordinated Activity to Expose Information Operations on Twitter. In *Proceedings of the ACM on Web Conference 2024*. 2530–2541.
[31] Mark EJ Newman. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46, 5 (2005), 323–351.
[32] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. 2018. Deepinf: Social Influence Prediction with Deep Learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2110–2119.
[33] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. 2018. Deep state space models for time series forecasting. *Advances in neural information processing systems* 31 (2018).
[34] Marian-Andrei Rizoiu, Alexander Soen, Shidi Li, Pio Calderon, Leanne J Dong, Aditya Krishna Menon, and Lexing Xie. 2022. Interval-censored Hawkes processes. *Journal of Machine Learning Research* 23, 338 (2022), 1–84.
[35] Marian-Andrei Rizoiu, Lexing Xie, Scott Sanner, Manuel Cebrian, Honglin Yu, and Pascal Van Hentenryck. 2017. Expecting to be HIP: Hawkes Intensity Processes for Social Media Popularity. In *Proceedings of the 26th International Conference on World Wide Web*. 735–744.
[36] Dietram A Scheufele and Nicole M Krause. 2019. Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences* 116, 16 (2019), 7662–7669.
[37] Lin Tian, Xiuzhen Zhang, and Jey Han Lau. 2021. Rumour detection via zero-shot cross-lingual transfer learning. In *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part I 21*. Springer, 603–618.
[38] Lin Tian, Xiuzhen Zhang, and Jey Han Lau. 2023. Metatroll: Few-shot Detection of State-Sponsored Trolls with Transformer Adapters. In *Proceedings of the ACM Web Conference 2023*. 1743–1753.
[39] Lin Tian, Xiuzhen Jenny Zhang, and Jey Han Lau. 2022. DUCK: Rumour Detection on Social Media by Modelling User and Comment Propagation Networks. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4939–4949.
[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *Advances in Neural Information Processing Systems* 30 (2017).
[41] Jia Wang, Vincent W Zheng, Zemin Liu, and Kevin Chen-Chuan Chang. 2017. Topological recurrent neural network for diffusion prediction. In *2017 IEEE international conference on data mining (ICDM)*. IEEE, 475–484.
[42] Yongqing Wang, Huawei Shen, Shenghua Liu, Jinhua Gao, and Xueqi Cheng. 2017. Cascade dynamics modeling with attention-based recurrent neural network. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 2985–2991.
[43] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Advances in Neural Information Processing Systems* 34 (2021), 22419–22430.
[44] Siqi Wu, Marian-Andrei Rizoiu, and Lexing Xie. 2018. Beyond Views: Measuring and Predicting Engagement in Online Videos. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.
[45] Xovee Xu, Fan Zhou, Kunpeng Zhang, Siyuan Liu, and Goce Trajcevski. 2021. Casflow: Exploring hierarchical structures and propagation uncertainty for cascade prediction. *IEEE Transactions on Knowledge and Data Engineering* 35, 4

Lin Tian, Emily Booth, Francesco Bailo, Julian Droogan, and Marian-Andrei Rizoiu

(2021), 3484–3499.

[46] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019. Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. In *Companion Proceedings of the 2019 World Wide Web Conference*. 218–226.

[47] Rui Zhang, Christian Walder, Marian-Andrei Rizoiu, and Lexing Xie. 2019. Efficient Non-parametric Bayesian Hawkes Processes. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* (California). International Joint Conferences on Artificial Intelligence Organization, 4299–4305. https://doi.org/10.24963/ijcai.2019/597

[48] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11106–11115.

[49] Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. 2020. Transformer hawkes process. In *International Conference on Machine Learning*. PMLR, 11692–11702.

A Ethics Considerations

Our research exclusively uses publicly available Facebook data on CrowdTangle, adhering to the platform's terms of service and research guidelines. We implement strict data protection measures: all user identifiers are anonymized, personal information is excluded from our analysis, and we focus solely on aggregate engagement patterns. Our data collection and processing procedures have been reviewed and approved. We maintain data minimization principles, collecting only information necessary for our research objectives.

B Experimental Settings

To evaluate our proposed IC-Mamba🌓 model, we conducted systematic experiments with different parameter configurations. This section includes our experimental environment, hyperparameters, and optimization approach.

B.1 Experimental Environment

All experiments were conducted using PyTorch 2.0 on a GPU cluster with 4xNVIDIA A100 GPUs with 40GB memory. All reported figures are averaged across ten runs with different random seeds.

B.2 Hyper-parameters

Table 5 lists the hyperparameter ranges used in our experiments.

C Mathematical Notations and Definitions

Table 6 summarizes the notations used throughout the paper to describe social outbreak events, associated posts, and the engagement predictions. Each notation is accompanied by a brief explanation for clarity and ease of reference.

D Next-time Social Engagement Prediction

While overall engagement prediction provides valuable insights into model performance, the ability to predict engagement at future time points is crucial for real-time social media monitoring and intervention. This section focuses on models capable of generating predictions for upcoming engagement values at the next future time point. We evaluate Informer [48], Autoformer [43], Mamba [11], and IC-Mamba🌓 for this task, as these models are architecturally designed for next-point prediction.

We follow the same temporal set up for next-time social engagement prediction task, with 6-hour history data as the input. For this task, we set up three different temporal stages of post lifecycle,

Table 5: IC-Mamba🌓 Hyperparameters

Parameter	Range
Hidden Size (N)	[256, 1024]
Input Dimension (D)	[64, 128]
Number of IC-Mamba Blocks	[2, 8]
Embedding Dimension (d)	[128, 512]
Interval-Censor State Dim	[512, 2048]
Sequence Length (L)	[1024, 8192]
Batch Size	[16, 128]
Learning Rate	[1e-5, 1e-3]
Warmup Steps	[500, 1650]
Weight Decay	[0.005, 0.15]
Dropout Rate	[0.15, 0.3]
λ (Loss Weight)	[(0.1, 0.9),(0.4,0.6)]
Training Epochs	[50, 150]
Early Stopping Patience	[5, 20]
τ (RTE Parameter)	[5000, 20000]
Time Granularity	[5m, 1h]
Max Prediction Window	[7d, 28d]
β_1	[0.85, 0.95]
β_2	[0.995, 0.9999]
ε	[1e-9, 1e-7]
Gradient Clipping	[0.5, 5.0]

each representing different length of intervals and data availability scenarios. Early-stage predictions (within first hour) have limited historical data but require quick response to emerging trends. Mid-stage predictions (within first day) balance data availability with evolving engagement patterns. Late-stage predictions (within first week) have rich historical context but must account for long-term engagement dynamics.

Fixed-Window Prediction (6-Hour Input). Table 7 presents the RMSE scores for engagement prediction using a fixed 6-hour historical window. All models use exactly 6 hours of historical data regardless of when the next engagement occurs. We observe that IC-Mamba🌓 consistently achieves the lowest RMSE scores across all datasets, indicating strong performance in capturing short-term temporal patterns. However, it's noteworthy that baseline models like Autoformer and Informer also perform competitively, suggesting that the fixed-window approach provides sufficient context for short-term prediction. An interesting finding is that the performance gap between IC-Mamba🌓 and Mamba is relatively small in this setting.

Early-Stage Prediction (≤ 1 hour). In the early-stage prediction task, models forecast the next engagement within the first hour of a post's publication. As shown in Table 8, all models experience increased RMSE compared to the fixed-window prediction, reflecting the challenge of making accurate predictions with limited historical data (typically 2-9 data points). Notably, IC-Mamba🌓 achieves the lowest RMSE, but the performance gap between IC-Mamba🌓 and Mamba widens in this setting.

Table 6: Notation Table

Notation	Description
$\mathcal{E}$	A social outbreak event
$\mathcal{P}$	The set of associated posts for $\mathcal{E}$, $\mathcal{P} = \{p_1, p_2, \ldots, p_N\}$
p	A single post $p \in \mathcal{P}$
t_0	The original posting time of a post p
x	The textual content of a post p
u	User metadata associated with the post
O	The set of possible opinion classes
o	The opinion class of a post, $o \in O$
$\mathcal{P}_o$	The set of posts sharing opinion o, $\{p \in \mathcal{P} \mid o(p) = o\}$
H	Interval-censored engagement history of a post, $H = \{(t_1, e_1), \ldots, (t_m, e_m)\}$
t_j	Observation time for engagement measurement in H
e_j	d-dimensional vector capturing different types of engagement at time t_j
Δt_j	Interval between consecutive observations, $\Delta t_j = t_{j+1} - t_j$
τ_{obs}	Observation window duration (e.g., 1 day)
$H_{\tau_{obs}}(p)$	Initial engagement history within observation window, $\{(t, e) \in H \mid t_0 \leq t \leq t_0 + \tau_{obs}\}$
T	Prediction horizon (e.g., 28 days)
τ_{step}	Fixed time interval for predictions (e.g., 5 minutes)
K	Number of prediction points, $K = \lfloor T/\tau_{step} \rfloor$
$\hat{e}(t)$	Predicted engagement at time t
$\hat{e}_{total}$	Predicted total cumulative engagement over horizon T

Table 7: Engagement prediction results with fixed 6-hour historical window; RMSE scores reported; lower is better; best results in boldface. All models use exactly 6 hours of historical data regardless of when the next engagement occurs.

Model	Bushfire	Climate	Vaccination	CoVID	DiN
Informer [48]	0.208	0.215	0.203	0.211	0.248
Autoformer [43]	0.196	0.203	0.191	0.199	0.234
Mamba [11]	0.152	0.158	0.147	0.155	0.184
IC-Mamba	**0.144**	**0.150**	**0.139**	**0.147**	**0.175**

Table 8: Early-stage engagement prediction results (next interval ≤ 1 hour); RMSE scores reported; lower is better; best results in boldface.

Model	Bushfire	Climate	Vaccination	CoVID	DiN
Informer [48]	0.245	0.252	0.238	0.248	0.299
Autoformer [43]	0.231	0.238	0.225	0.234	0.284
Mamba [11]	0.188	0.194	0.193	0.201	0.265
IC-Mamba	**0.169**	**0.175**	**0.164**	**0.172**	**0.235**

Another observation is that the baseline models, Informer and Autoformer, show a heavy drop in performance during early-stage predictions. Additionally, the DiN dataset shows higher RMSE scores across all models, indicating that early-stage prediction is particularly challenging for post related to disinformation.

Table 9: Mid-stage engagement prediction results (next interval ≤24 hours); RMSE scores reported; lower is better; best results in boldface.

Model	Bushfire	Climate	Vaccination	CoVID	DiN
Informer [48]	0.198	0.205	0.193	0.201	0.235
Autoformer [43]	0.187	0.194	0.182	0.190	0.223
Mamba [11]	0.142	0.148	0.137	0.145	0.172
IC-Mamba	**0.135**	**0.141**	**0.130**	**0.138**	**0.164**

Table 10: Late-stage engagement prediction results (next interval ≤1 week); RMSE scores reported; lower is better; best results in boldface.

Model	Bushfire	Climate	Vaccination	CoVID	DiN
Informer [48]	0.165	0.137	0.152	0.168	0.175
Autoformer [43]	0.156	0.128	0.163	0.159	0.168
Mamba [11]	0.118	0.123	0.114	0.121	0.144
IC-Mamba	**0.112**	**0.117**	**0.108**	**0.115**	**0.137**

Mid-Stage Prediction (≤24 hours). In the mid-stage predictions, with more historical data available, all models show improved RMSE scores (Table 9). The performance gap between the models becomes smaller, indicating that the availability of additional data helps all models make better predictions. IC-Mamba continues to outperform the baselines.

An interesting finding is that the performance on Vaccination theme shows a significant reduction in RMSE for all models in the mid-stage prediction. This may imply that engagement patterns for vaccination-related content become more predictable within the first day, possibly due to sustained public interest and consistent interaction patterns.

Late-Stage Prediction (≤1 week). For late-stage predictions, with extensive historical data (up to one week), all models achieve their best RMSE scores (Table 10). The performance differences between models are less pronounced, though IC-Mamba still holds a slight advantage.

We found that the Climate theme shows relatively low RMSE scores across all models in late-stage prediction. This could reflect consistent engagement patterns over longer periods for climate-related content, perhaps due to sustained public interest and ongoing discussions.