CasUGC: Aligning User-Generated Comment Evolution with Cascade Dynamics for Popularity Prediction

Yifan Meng[1] , **Xigang Sun**[1] , **Anran Zhang**[1] , **Jiaqi Jiang**[1] and **Jiahui Jin**[1*]

[1]School of Computer Science and Engineering, Southeast University, China

{mengyifan, xigangsun, arzhang, jiaqijiang, jjin}@seu.edu.cn

Abstract

Accurately predicting the popularity of information cascades facilitates the development of social media network applications. During the cascade propagation process, user-generated comments continually evolve, thereby substantially affecting overall information popularity. However, existing methods primarily focus on learning structural–temporal cascade dynamics while neglecting the modeling of user-generated comment evolution, leading to suboptimal predictions. In this paper, we propose a novel framework, CasUGC, that aligns Cascade dynamics with User-Generated Comment evolution for popularity prediction. Specifically, we develop a dual-granularity alignment strategy that bridges the representation gap between comment semantics and structural–temporal dynamics at both the user and cascade levels. Building on this alignment, we further design a dynamic cascade feature generation module to produce temporally enhanced cascade embeddings. To further address the inherent uncertainties in both alignment and prediction, we introduce two diffusion-based components: a cascade-level diffusion network that learns a latent distribution mapping between semantic and structural–temporal representations, and a second diffusion process that captures the stochasticity of propagation dynamics for robust prediction. Extensive experiments conducted on four real-world social media datasets demonstrate the superiority of CasUGC over state-of-the-art methods.

1 Introduction

Cascade popularity prediction aims to estimate the number of comments or reposts influenced by users' content preferences based on early-stage observations [Li *et al.*, 2017; Xu *et al.*, 2021]. It plays a fundamental role in applications such as public opinion monitoring, viral marketing, and recommendations on online social media platforms [Zhou *et al.*, 2021]. Accurately forecasting whether content will achieve widespread diffusion benefits both platform regulation and business decision-making.

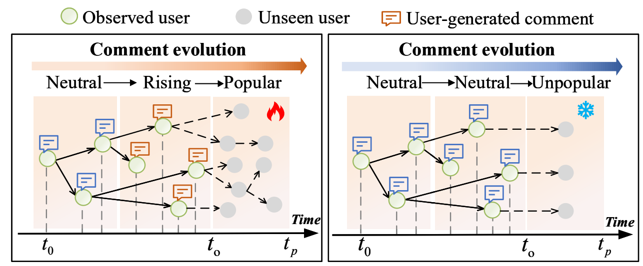

Figure 1: A case of information cascade with comment evolution. A user posts one comment at time t_0. During observation $[t_0, t_o]$, other users participate in the repost or comment process. Observed cascades (a) and (b) appear similar but lead to different future popularity from t_o to t_p due to comment evolution.

Most current studies on cascade popularity prediction primarily focus on modeling the structural and temporal dynamics of the user reposting or resharing process using spatial-temporal neural networks [Lu *et al.*, 2023; Zhou *et al.*, 2024; Jiao *et al.*, 2024; Cheng *et al.*, 2024b; Jing *et al.*, 2025; Bao *et al.*, 2024]. Another approach involves using fusion networks to learn representations of multimodal content for predicting popularity [Zhang *et al.*, 2018; Hsu *et al.*, 2019; Chen *et al.*, 2023; Zhong *et al.*, 2024; Xu *et al.*, 2025b], without considering structural-temporal dynamics of cascades. However, in real-world social platforms, cascade popularity is reciprocally influenced by structural and temporal characteristics and the evolution of user-generated comments. For instance, users often refer to previous content when reposting or commenting, forming a comment evolution process. As Figure 1 shows, under limited early-stage observations, cascades with similar structures evolve into different popularity levels due to the influence of user-generated comment evolution. Introducing comment evolution can enrich the observed propagation signals and thus improve popularity prediction accuracy. Moreover, some uncertain factors, such as sudden trending events or rapid rumor debunking, can cause cascades to either burst or fade. *Overall, existing methods overlook the interplay of comment evolution and cascade dynamics with propagation uncertainty, which is crucial to understanding how cascades gain different popularity outcomes.*

*Corresponding Author

Motivated by these, we investigate the cascade popularity prediction task with user-generated comments in this paper. Nevertheless, bridging the gap between structural-temporal dynamics and user-generated comment evolution influence in cascades is non-trivial. Specifically, achieving this goal is hindered by two significant challenges.

First, inconsistent representation spaces in cascades. User-generated comment evolution reflects semantic and contextual influence in a high-dimensional semantic space. In contrast, structural-temporal dynamics capture cascade propagation along network topology and time in a topological space. Bridging these heterogeneous spaces for effective interaction is challenging. Existing studies [Li *et al.*, 2017; Li *et al.*, 2018; Lu *et al.*, 2023; Xu *et al.*, 2025b] often strengthen structural-temporal modeling or incorporate content features in isolation. They overlook the complex, nonlinear interplay between what is shared and how, when, and where it propagates.

Second, inherent uncertainty in alignment and prediction. Uncertainty arises at two critical stages of cascade modeling. Alignment uncertainty means the mapping between a comment and its potential propagation pattern is not deterministic. Similar comments may lead to different cascade trajectories under varying circumstances, making it difficult to learn a one-to-one correspondence between the semantic and structural-temporal spaces. Prediction uncertainty persists even after successful representation alignment, as cascade evolution is inherently stochastic. The ultimate popularity is affected by unpredictable factors, such as real-world events or platform interventions. Existing approaches [Cao *et al.*, 2017; Lu *et al.*, 2023; Cheng *et al.*, 2024b; Jing *et al.*, 2025] mostly assume deterministic mappings or fail to model alignment uncertainty explicitly.

To overcome these challenges, we propose a novel framework, CasUGC, that aligns Cascade dynamics with User-Generated Comment evolution for popularity prediction. First, we design a dual-granularity alignment module to bridge the inconsistent representation spaces in cascades. It integrates user-level contrastive learning and cascade-level feature alignment on separately embedded spaces of the cascade and comment graphs. Based on this, we further develop a dynamic cascade feature generation module to obtain temporally enhanced cascade embeddings. Second, to capture the inherent uncertainty in alignment and prediction, we introduce a two-stage diffusion generation strategy. Specifically, a conditional diffusion network captures latent distribution mapping between semantic and structural–temporal spaces during cascade-level alignment. Another diffusion process captures the uncertainty of propagation bursts within dynamic feature generation. We collect four real-world datasets containing rich user-generated comments. Extensive experiments demonstrate that the proposed CasUGC achieves state-of-the-art on all datasets. Our contributions are:

- We highlight the critical yet underexplored role of user-generated comment evolution in cascade popularity prediction. In light of this, we propose CasUGC, which explicitly models the interplay between comment evolution and cascade dynamics.
- We introduce a dual-granularity alignment module to unify comment evolution and structural–temporal spaces. A dynamic cascade feature generation module is designed to obtain temporally enhanced cascade embeddings. We further employ two cascade diffusion models to capture alignment and prediction uncertainty.
- We construct four real-world social media datasets containing rich user-generated comments and conduct comprehensive experiments. Results demonstrate that CasUGC consistently outperforms state-of-the-art baselines, validating the effectiveness of modeling comment evolution in cascade popularity prediction.

2 Preliminaries

Comment evolution characterizes the dynamic process in which user-generated comments accumulate and shift over the lifetime of an information cascade. It encodes the semantic trajectory of the conversation—how earlier comments and the underlying cascade context influence each new remark. In this section, we revisit the definition of the cascade and popularity prediction by incorporating comments.

Cascade with comments: An *information cascade* is a propagation trace that records how information spreads among users through actions such as reposts or comments. Formally, a cascade with K users is represented as a set of tuples

$$\mathcal{C} = \{(u_j, u_i, t_i, c_i)\}_{i,j<K}, \tag{1}$$

where u_i denotes the user participating in the cascade at time t_i via performing an action in response to u_j, and c_i represents the textual comment generated by u_i. Let $\mathcal{C}(t_o) = \{(u_j, u_i, t_i, c_i) | 0 \leq i < k\}$ denote the observed cascade with k users up to time t_o.

Cascade graph: Given an observed cascade $\mathcal{C}(t_o) = \{(u_j, u_i, t_i, c_i)\}$, and $t_i \leq t_o$, the corresponding cascade graph denote as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V} = \{u_i \mid 0 \leq i < k\}$ denotes all involved users and the edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ captures the propagation actions among users.

Cascade comment graph: Similar to cascade graph, given an observed cascade $\mathcal{C}(t_o)$, the corresponding comment graph denote as $\mathcal{G}^c = (\mathcal{V}^c, \mathcal{E}^c)$, the edge set $\mathcal{E}^c$ is constructed as follows:

$$\begin{aligned} \mathcal{E}^c = \{(c_j, c_i) \mid \ if \ (u_j, u_i) \in \mathcal{C}(t_o)\} \\ \cup \{(c_i, c_{i+1}) \mid 0 \leq i < k\}, \end{aligned} \tag{2}$$

where the propagation edges $\{(c_j, c_i)\}$ denotes the propagation structure and the edges of temporal order $\{(c_i, c_{i+1})\}$ represent the cumulative impact of comments evolution.

Cascade popularity prediction problem. Given a cascade $\mathcal{C}(t_o)$ within observation time interval $[0, t_o]$, the goal is to learn a prediction function $f_\Theta(\cdot)$ parameterized by Θ to estimate the future cascade size in the subsequent window $[t_o, t_p]$ by exploiting the comment signals already available:

$$\Delta\hat{y} = f_\Theta\big(\mathcal{C}(t_o)\big), \tag{3}$$

where $\Delta\hat{y}$ denotes the predicted increment and the ground-truth is $\Delta y = |\mathcal{C}(t_p)| - |\mathcal{C}(t_o)|$.

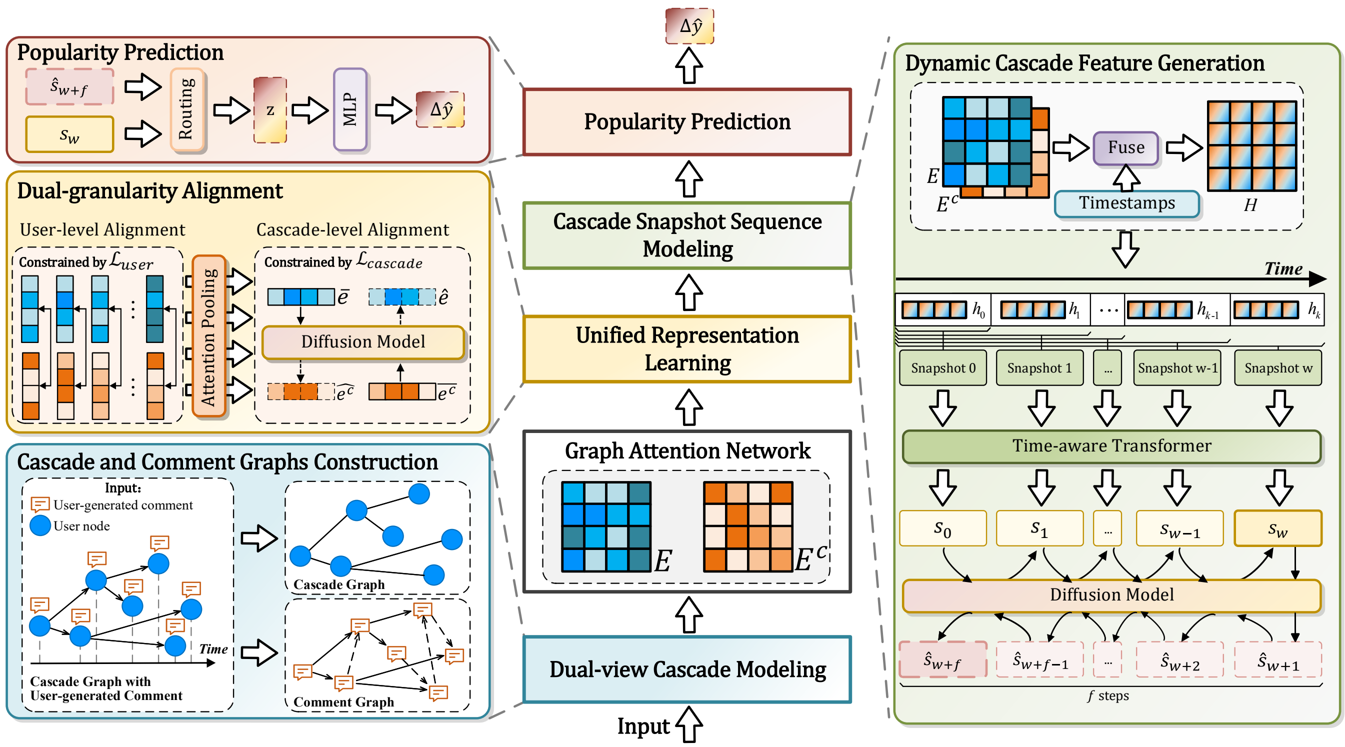

Figure 2: The overview framework of our proposed CasUGC. It consists of a cascade and comment graphs construction module to learn cascade structural and comment embedding, a dual-granularity alignment module to align two heterogeneous spaces, a dynamic cascade feature generation module to generate future cascade representation, and a prediction module to give incremental popularity.

3 Methodology

As shown in Figure 2, we introduce **CasUGC**, a framework for cascade popularity prediction that aligns comment evolution with cascade dynamics. CasUGC first constructs cascade and comment graphs to obtain topological and semantic embeddings. To bridge heterogeneous representation spaces, we propose a dual-granularity alignment module that leverages a conditional diffusion process to explicitly capture alignment uncertainty between semantics and structure. Furthermore, we design a dynamic cascade feature generation module to model the stochastic evolution of snapshots, incorporating prediction uncertainty into the final popularity estimation.

3.1 Dual-view Cascade Modeling

Information cascade propagation is jointly influenced by cascade dynamics and user-generated comment evolution, and thus calls for a unified representation space. However, directly injecting comment features into user embeddings risks entangling heterogeneous information sources, since structural–temporal propagation patterns and semantic comment evolution follow different ways. To address this, we design a cascade graph and comment graph construction strategy in a dual-view setting, which models the cascade dynamics and comment evolution as two complementary graphs derived from the same observed cascade.

To capture the cascade propagation pattern among users in reposting or commenting, we introduce a cascade graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from the observed cascade $\mathcal{C}(t_o) = \{(u_j, u_i, t_i, c_i)|0 \le i < k\}$. We then extract cascade embedding $X = \{x_i\} \in \mathbb{R}^{k\times d}$ for each node using the spectral graph wavelet method [Donnat *et al.*, 2018] with normalized temporal weights $\{(t_o - t_i)/t_o|t_i \le t_o\}$, where d is the dimension of the cascade embedding. The embedding encodes the user-individual structural-temporal representations in the topological space.

To capture the evolution of user-generated comments, we introduce a comment graph $\mathcal{G}^c = (\mathcal{V}^c, \mathcal{E}^c)$. Considering cumulative semantic influence, we add directed edges $\{(c_i, c_{i+1})\}$ between temporally adjacent comment nodes, reflecting that users observe prior comments before posting. After that, we generate comment embedding X^c for each node by feeding the user-generated comment c_i into a pretrained Sentence-BERT[Reimers and Gurevych, 2019]. This representation $X^c = \{x^c_i\} \in \mathbb{R}^{k\times d_c}$ captures semantic and contextual information, where d_c is the dimension of the comment embedding space.

We then apply Graph Attention Networks (GAT) [Velickovic *et al.*, 2017] to process $\mathcal{G}$ and $\mathcal{G}^c$ with pre-representation X and X^c, capturing neighboring dependencies via a powerful attention mechanism to obtain the cascade and comment node embeddings: $E = [e_0, \dots, e_k], E^c = [e^c_0, \dots, e^c_k] \in \mathbb{R}^{k\times d_h}$, where d_h denotes the dimension of node embeddings. These embeddings are leveraged for subsequent unified representation learning.

3.2 Unified Representation Learning

While cascade and comment graphs provide complementary insights, their embedding spaces remain intrinsically inconsistent. Furthermore, the non-deterministic mapping between semantic content and structural-temporal patterns introduces significant alignment uncertainty. To mitigate these issues, we propose a dual-granularity alignment mechanism that harmonizes representations at both user and cascade levels, effectively capturing these underlying stochastic correlations.

User-level Alignment via Contrastive Learning

At the user level, we aim to enforce consistency between a user's structural-temporal embedding e_i and comment embedding e_i^c. We adopt a contrastive alignment objective [Radford *et al.*, 2021] that pulls matched user–comment pairs together while separating non-matching pairs within a minibatch. Formally, the user-level alignment loss is defined as:

$$\mathcal{L}_{\text{user}} = -\frac{1}{2(k+1)} \sum_{i=0}^{k} \left[\log \frac{\exp(\text{sim}(e_i, e_i^c)/\tau)}{\sum\limits_{j \neq i} \exp(\text{sim}(e_i, e_j^c)/\tau)} + \log \frac{\exp(\text{sim}(e_i^c, e_i)/\tau)}{\sum\limits_{j \neq i} \exp(\text{sim}(e_i^c, e_j)/\tau)} \right], \quad (4)$$

where $\text{sim}(\cdot,\cdot)$ is the cosine similarity, and $\tau > 0$ is a temperature hyper-parameter. Minimizing this loss fosters a unified latent space in which user nodes are mutually reinforced by their textual context, preventing the dissociation of the propagation structure from its underlying semantic evolutions.

Cascade-level Alignment via Conditional Diffusion

While user-level contrastive learning improves local representation consistency, it fails to capture the global correlation between cascade dynamics and comment evolution. Establishing this global alignment is non-trivial due to inherent alignment uncertainty. In other words, the mapping between topological space and semantic context is often stochastic and one-to-many, making global alignment fundamentally ambiguous. To address this, we introduce a cascade-level alignment mechanism utilizing a conditional diffusion model.

The diffusion model [Ho *et al.*, 2020] learns a probabilistic distribution over semantic evolutions conditioned on structural–temporal patterns (and vice versa). It resolves alignment uncertainty in a generative manner. Unlike VAEs or GANs, diffusion models can capture the complex and multimodal distributions, making them suited to the stochastic mapping between structural-temporal and semantic spaces.

Let $\bar{e}$ and $\bar{e^c}$ denote cascade-level representations aggregated from the cascade and comment graphs, respectively, via an attention pooling operator $\text{Agg}(\cdot)$: $\bar{e} = \text{Agg}(E)$, $\quad \bar{e^c} = \text{Agg}(E^c)$. We implement a bidirectional diffusion process (see Appendix A) to achieve mutual alignment. Taking the structural perspective as an example, we define $\bar{e}$ as the clean state $\bar{e}_0$ and construct a forward Markov chain by progressively injecting Gaussian noise over T steps. The reverse process then learns to reconstruct the structural-temporal signal $\bar{e}_0$ conditioned on the semantic context $\bar{e^c}$. A symmetric process is simultaneously applied to reconstruct $\bar{e^c}$ conditioned on $\bar{e}$, ensuring bidirectional alignment.

This bidirectional framework is formalized as:

$$\widehat{e} = \text{ConDiff}(\bar{e} \mid \bar{e^c}), \qquad \widehat{e^c} = \text{ConDiff}(\bar{e^c} \mid \bar{e}), \quad (5)$$

yielding reconstructed embeddings $\widehat{e}$ and $\widehat{e^c}$ under counterpart conditions. The model is trained to minimize two reconstruction errors, defined as:

$$\mathcal{L}_{\text{rec}}^{u} = \left\| \widehat{e} - \bar{e} \right\|_2^2, \qquad \mathcal{L}_{\text{rec}}^{c} = \left\| \widehat{e^c} - \bar{e^c} \right\|_2^2. \quad (6)$$

To further enforce invariance under noise perturbations, we introduce a cascade-level matching loss that aligns the two reconstructed representations:

$$\mathcal{L}_{\text{match}} = \left\| \widehat{e} - \widehat{e^c} \right\|_2^2. \quad (7)$$

Thus, the cascade-level alignment loss is:

$$\mathcal{L}_{\text{cascade}} = \mathcal{L}_{\text{rec}}^{u} + \mathcal{L}_{\text{rec}}^{c} + \mathcal{L}_{\text{match}}, \quad (8)$$

which jointly (i) denoises each type of representation via reconstruction and (ii) couples the two representations by minimizing the distance between their diffusion-generated counterparts. Unlike deterministic approaches, the proposed method enables stable training and achieves a diverse yet coherent global alignment under the inherent uncertainty between structural-temporal patterns and semantic evolutions.

3.3 Cascade Snapshot Sequence Modeling

We employ snapshot encoding on aligned cascade representations to model the dynamic accumulation of popularity in cascades with comments. Relying solely on observed data neglects the prediction uncertainty that similar early-stage dynamics can evolve into vastly different popularity. To explicitly model the inherent stochasticity of future cascade growth, we propose the *Dynamic Cascade Feature Generation* module, which leverages a generative diffusion mechanism to produce uncertainty-aware snapshot sequence embeddings for future evolution.

Cascade Snapshot Encoding

Given the observed time window $[0, t_o]$, the dynamic cascade feature diffusion generation module partitions it into w equal-length intervals delimited by timestamps $t_0, t_1, \ldots, t_w$, where $t_0 = 0$, $t_w = t_o$. The resulting sequence of cascade snapshots $\mathcal{C}(t_0) \to \mathcal{C}(t_1) \to \cdots \to \mathcal{C}(t_o)$ succinctly depicts the cascade dynamics.

To encode each snapshot $\mathcal{C}(t_n)$, we first integrate the aligned user and comment representations of each participant. For every tuple $(u_j, u_i, t_i, c_i) \in \mathcal{C}(t_n)$, we project the timestamp t_i into the same latent dimension as e_i and e_i^c via multi-layer perceptron and then fuse the three signals: $h_i = \text{LN}\big(e_i + e_i^c + \text{MLP}(t_i)\big)$, where $\text{LN}(\cdot)$ denotes layer normalization for stable training.

The snapshot representation S_n is produced by aggregating all tuple encodings h_i within $\mathcal{C}(t_n)$. To inject snapshot-level temporal context, we utilize an MLP to generate positional encodings for n, and construct the sequence $H_n = \{h_i + \text{MLP}(n) \mid t_i < t_n\}$. Feeding H_n into a Transformer encoder and extracting the snapshot representation:

$$S_n = \text{TransformerEncoder}(H_n), n = 0, 1, \ldots, w. \quad (9)$$

Stacking these representations across all windows gives the temporally-aware sequence $[S_0, S_1, \ldots, S_w]$, which serves as the foundation for modeling cascade evolution.

Diffusion-based Snapshot Sequence Generation

Cascade popularity prediction is inherently challenging due to prediction uncertainty, such as sudden bursts driven by trending events or rapid fading. Therefore, we apply a diffusion model to the transition between adjacent snapshots, capturing short-term prediction uncertainty. Formally, for every consecutive pair $(\mathcal{C}(t_n), \mathcal{C}(t_{n+1}))$, we learn a conditional diffusion process $\widehat{S}_{n+1} = \mathrm{ConDiff}(S_{n+1} \mid S_n)$, and minimize the reconstruction error $\mathcal{L}^i_{\text{rec}} = \left\|\widehat{S}_{n+1} - S_{n+1}\right\|^2_2$. The training objective averages this loss over all observed transitions:

$$\mathcal{L}_{\text{gen}} = \frac{1}{w}\sum_{i=1}^{w}\mathcal{L}^i_{\text{rec}}. \tag{10}$$

After that, we recursively unroll the future snapshots: starting from the last observed snapshot S_w, we iteratively denoise to produce $\widehat{S}_{w+1},\ \widehat{S}_{w+2},\ \ldots,\ \widehat{S}_{w+f}$, each step using the previously generated snapshot as the new condition. The *prev-to-next* conditioning enforces temporal coherence, while stochasticity from diffusion preserves uncertainty in comment evolution and cascade dynamics, yielding consistent and uncertainty-aware future cascade representations.

3.4 Popularity Prediction

In the final prediction stage, we aim to infer future popularity by integrating two distinct propagation signals. First, the last observed snapshot feature S_w encapsulates the cumulative historical context of the cascade, thereby serving as a robust indicator of the stable, long-term trend. Second, the generative future snapshot feature $\widehat{S}_{w+f}$, derived via the stochastic diffusion process, explicitly models prediction uncertainty, capturing potential bursts and transient fluctuations that deterministic history cannot reveal.

To adaptively balance short-term and long-term information, we design a routing mechanism in the predictor head that dynamically fuses the generated future feature $\widehat{S}_{w+f}$ and the last observed snapshot S_w. Specifically, a sigmoid function computes a routing score α from these two features. We learn a weighted combination of them:

$$z = \mathrm{Routing}(\widehat{S}_{w+f}, S_w) = \alpha S_w + (1-\alpha)\widehat{S}_{w+f}. \tag{11}$$

The resulting fused representation z captures both the stable long-term trend and the transient short-term fluctuation of the cascade, providing a balanced feature for prediction.

This fused representation is input into a prediction head (MLP): $\Delta\hat{y} = \mathrm{MLP}(z)$. The final training objective of our model integrates all previously proposed loss functions, formulated as follows:

$$\mathcal{L} = \|\Delta\hat{y} - \Delta y\|^2 + \lambda_1\mathcal{L}_{\text{user}} + \lambda_2\mathcal{L}_{\text{cascade}} + \lambda_3\mathcal{L}_{\text{gen}}, \tag{12}$$

where Δy is the true incremental cascade popularity.

4 Experiments

We compare CasUGC with the state-of-the-art baselines for cascade popularity prediction on four real-world datasets, aiming to answer the following research questions: **Q1)** What advantages does the CasUGC achieve compared to other methods for the information cascade popularity prediction? **Q2)** Can each component of CasUGC contribute to improving the cascade popularity prediction? **Q3)** How sensitive is the CasUGC to the different settings for important hyperparameters? **Q4)** How well does CasUGC explain its prediction capabilities in the context of user-generated comments?

Dataset	WeiboMSF	WeiboBE	LC-XHS	LC-TB
# Cascades	10,203	4,568	3,116	4,012
# Users	880,707	373,139	313,920	385,487
Avg. popularity	256	83	110	96
Min. popularity	50	20	20	20
Max. popularity	63,600	500	3,462	5,000

Table 1: Descriptive statistics of four social media datasets.

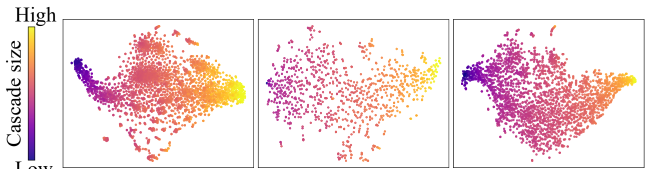

Figure 3: Visualization of user-generated comment embeddings in cascades after dimensionality reduction using t-SNE.

4.1 Experimental Setup

Datasets. Most available cascade datasets only provide structural-temporal information and lack the accompanying user-generated comments. We construct four social media datasets with rich user-generated comments: WeiboMSF[Wu *et al.*, 2022], WeiboBE[1], LC-XHS, and LC-TB. The statistical details are presented in Table 1.

The WeiboMSF and WeiboBE datasets are derived from publicly available sources and processed in accordance with our experimental requirements. The LC-XHS and LC-TB datasets are collected from lifestyle communities in Xiaohongshu and Tieba. We select 30 high-activity topics from trending tags and retrieve posts with complete comment. Abnormal records are filtered out to ensure data quality.

As shown in Figure 3, the dataset indicates that the semantics of user-generated comments play a crucial role in influencing the scale of the cascade. This observation motivates us to incorporate comment evolution into popularity prediction.

Metrics and Baselines. Three metrics are employed in the experiments: Mean Squared Logarithmic Error (MSLE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). We select eleven SOTA baselines for comparison: (1) *Sequence modeling methods*: DeepCas[Li *et al.*, 2017], DeepHawkes[Cao *et al.*, 2017], DCGT[Li *et al.*, 2018], (2) *Propagation structure modeling methods*: CasTformer[Sun *et al.*, 2023b], CasFlow[Xu *et al.*, 2021], MINDS[Jiao *et al.*, 2024], (3) *Continuous-time and uncertainty-aware methods*: CTCP[Lu *et al.*, 2023], CasDO[Cheng *et al.*, 2024b], CasFT[Jing *et al.*, 2025], (4) *Multimodal prediction methods*: SKAPP[Xu *et al.*, 2025b], and SMTPD[Xu *et al.*, 2025c].

[1] https://magic-group-buaa.github.io/IJCAI25/pages/task2.html

Model	WeiboMSF			WeiboBE			LC-XHS			LC-TB		
	MSLE	MAPE	R^2	MSLE	MAPE	R^2	MSLE	MAPE	R^2	MSLE	MAPE	R^2
DeepCas	4.169	0.513	0.181	3.754	0.588	0.406	4.678	0.595	-0.028	4.726	0.728	-0.035
DeepHawks	4.279	0.455	0.119	3.323	0.399	0.560	4.796	0.561	0.094	4.275	0.653	-0.007
DCGT	4.253	0.494	0.135	3.254	0.314	0.321	4.588	0.552	0.069	4.124	0.613	0.013
CasTformer	3.733	0.436	0.129	2.694	0.399	0.724	4.065	0.460	0.147	3.960	0.619	0.076
CasFlow	3.431	0.292	0.292	2.459	0.218	0.699	3.834	0.347	0.188	2.960	0.389	0.352
MINDS	OOM	OOM	OOM	2.354	0.208	0.714	3.279	0.332	0.136	3.361	0.481	0.249
CTCP	3.167	0.308	0.277	1.955	0.187	**0.884**	4.964	0.348	0.086	3.460	0.527	0.152
CasDO	3.232	0.299	0.126	1.948	0.189	0.803	3.451	0.372	0.124	2.629	0.381	0.344
CasFT	3.253	0.357	0.313	2.233	0.223	0.763	**2.459**	0.307	0.221	2.485	0.458	**0.409**
SKAPP	3.843	0.453	0.107	2.644	0.566	0.634	4.074	0.458	0.157	3.886	0.592	0.118
SMTPD	3.787	0.396	0.176	2.588	0.263	0.725	3.173	0.368	0.189	3.735	0.614	0.113
CasUGC	**2.978**	**0.241**	**0.361**	**1.919**	**0.148**	0.771	2.533	**0.257**	**0.237**	**2.465**	**0.296**	0.352

Table 2: Performance comparison of CasUGC and other baseline models on WeiboMSF, WeiboBE, LC-XHS, and LC-TB. The best and second-best performances are highlighted in bold and underlined, respectively. OOM indicates the out-of-memory error.

Implementation details. The model architecture uses the following hyperparameters: 256-dimensional structural embedding, 768-dimensional comment embedding, and 128-dimensional structural and comment embeddings encoded through a 2-layer GAT. The diffusion model is configured with 64 timesteps.

4.2 Overall Performance Comparisons (RQ1)

Table 2 presents the overall performance comparison of our proposed CasUGC and other baseline methods in terms of MSLE, MAPE, and R^2 across four datasets. On the WeiboMSF dataset, CasUGC consistently outperforms all state-of-the-art methods, achieving improvements of 5.9%, 17.2%, and 15.0% on MSLE, MAPE, and R^2, respectively. Similarly, on the other three datasets, CasUGC achieves notable gains on most evaluation metrics.

We observe distinct performance patterns among the baselines: (1) Sequence modeling methods (DeepCas, DeepHawkes, DCGT) exhibit the weakest performance due to their limited capacity to capture complex structural-semantic interactions. (2) Multimodal methods (SKAPP, SMTPD), while incorporating content, fail to model propagation dynamics adequately. (3) Propagation structure modeling methods (CasFlow, MINDS) and Uncertainty-aware models (CTCP, CasDO, CasFT) achieve better results but still lag behind CasUGC. The superiority of CasUGC validates the effectiveness of dual-granularity alignment in unifying comment evolution with cascade dynamics, and the advantage of diffusion-based modeling in capturing alignment and prediction uncertainties.

4.3 Effectiveness of Component (RQ2)

We evaluate six variants of CasUGC to assess each component's contribution. The variants are defined as follows: (1) *w/o comment*: removing comment embeddings; (2) *w/o user align*: removing user-level alignment; (3) *w/o cas align*: removing cascade-level alignment; (4) *w/o struct*: removing structural embeddings; (5) *w/o temporal*: removing temporal encoding; and (6) *w/o diffusion*: removing diffusion-based uncertainty modeling in snapshot sequence generation.

Model	WeiboMSF			LC-XHS		
	MSLE	MAPE	R^2	MSLE	MAPE	R^2
w/o comment	3.565	0.281	0.207	2.732	0.277	0.185
w/o user align	3.016	0.233	0.247	2.738	0.274	0.182
w/o cas align	3.283	0.241	0.225	2.852	0.284	0.167
w/o struct	3.357	0.274	0.214	3.094	0.295	0.151
w/o temporal	3.328	0.265	0.285	2.982	0.291	0.169
w/o diffusion	3.167	0.254	0.315	2.652	0.267	0.202
CasUGC	**2.978**	**0.241**	**0.361**	**2.533**	**0.257**	**0.237**

Table 3: Ablation Study on WeiboMSF and LC-XHS datasets.

As shown in Table 3, *w/o comment* confirms that comment semantics are critical for cascade popularity; *w/o user align* and *w/o cas align* validate the dual-granularity alignment as essential for bridging heterogeneous representation spaces; *w/o struct* and *w/o temporal* reveal severe deterioration, confirming that structural and temporal signals form the foundation of cascade modeling; *w/o diffusion* yields notably higher MSLE and lower R^2, demonstrating the diffusion module's unique role in uncertainty-aware representation learning.

4.4 Hyperparameter Analysis (RQ3)

We investigate the sensitivity of CasUGC to three critical hyperparameters: the number of cascade snapshots w, the generated future snapshot steps f, and the diffusion timesteps h. Figure 4 depicts the performance trajectories on MSLE and R^2, revealing a consistent trend where performance peaks at intermediate values before degrading.

We observe that increasing w improves performance when $w < 10$, while larger values introduce instability due to redundant temporal signals. Performance also benefits from increasing f up to 3, beyond which additional steps may accumulate generation noise. Similarly, increasing h enhances performance below 64, whereas excessive diffusion steps de-

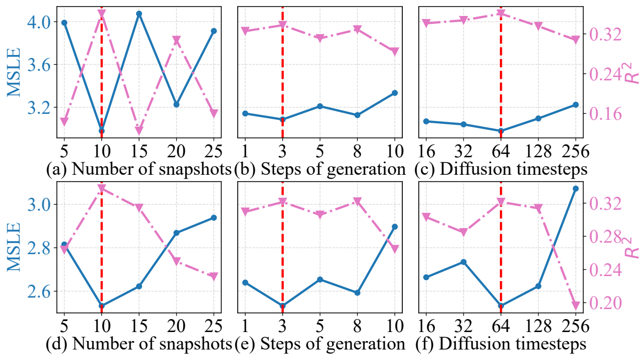

Figure 4: Hyperparameter analysis on two datasets. Subfigures (a)-(c) are based on the WeiboMSF dataset, while subfigures (d)-(f) are based on the LC-XHS dataset.

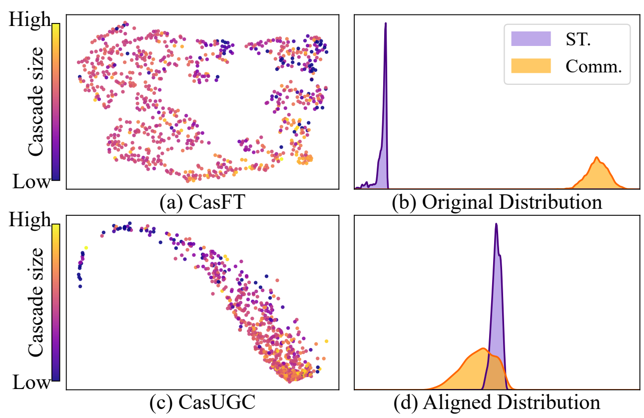

Figure 5: Visualization on the WeiboMSF dataset. Subfigures (a) and (c) are visualizations of learned cascade representations using t-SNE. Subfigures (b) and (d) present the KDE distributions of the structural-temporal (ST.) and comment (Comm.) patterns before and after alignment, respectively.

grade accuracy by amplifying stochastic noise and hindering convergence. Based on these observations, we set $w = 10$, $f = 3$, and $h = 64$ in all experiments, achieving better prediction accuracy.

4.5 Visualization Analysis (RQ4)

Figure 5 provides an intuitive visualization of learned cascade representations on the WeiboMSF dataset. As shown in subfigures (a) and (c), the embedding space of CasFT appears relatively scattered, with less distinct clustering and weaker differentiation across cascade sizes. In contrast, CasUGC learns a more structured manifold with a smooth color gradient across cascade sizes, demonstrating its superiority in obtaining discriminative cascade features. Subfigures (b) and (d) further illustrate the distributional alignment between the structural-temporal and comment patterns: they exhibit a clear distributional shift before alignment, whereas under the alignment mechanism of CasUGC, their KDE distributions largely overlap, indicating that semantic and structural-temporal representations are effectively unified.

5 Related Work

Traditional approaches. Early prediction methods mainly relied on hand-crafted features [Suh *et al.*, 2010; Bandari *et al.*, 2012], user attributes [Suh *et al.*, 2010; Ma *et al.*, 2013], structural cues [Cheng *et al.*, 2014; Bao *et al.*, 2013], and early temporal signals [Szabó and Huberman, 2010; Pinto *et al.*, 2013], combined with classical regressors or classifiers. Representative statistical generative models include different point process [Wang *et al.*, 2013; Shen *et al.*, 2014; Rizoiu *et al.*, 2017] and other models (e.g., SEISMIC [Zhao *et al.*, 2015], survival analysis [Yu *et al.*, 2015]. These traditional approaches rely on predefined features or low-dimensional parameterizations and struggle to capture complex interactions in cascades.

Structural-temporal deep learning approaches. Deep learning methods [Li *et al.*, 2017; Cao *et al.*, 2017; Wu *et al.*, 2022; Sun *et al.*, 2023a; Zhou *et al.*, 2024; Xu *et al.*, 2025a] for cascade popularity prediction learn end-to-end representations of propagation structure and temporal dynamics. More recent methods emphasize temporal encoding and generative modeling: CasDO[Cheng *et al.*, 2024b] models irregular temporal dynamics and uncertainty via probabilistic diffusion and neural ODEs, and CasFT[Jing *et al.*, 2025] leverages dynamic cues–driven diffusion models to capture future trend signals. Although these approaches can effectively capture structural–temporal interactions, they ignore the critical influence of the comment evolution on popularity acquisition.

Multi-media popularity prediction. Multi-media popularity prediction aims to forecast the potential reach of social media content by leveraging multimodal signals such as visual appearance, textual context, and social interactions, which has become increasingly important with the rise of image- and video-centric platforms [Zhang *et al.*, 2018; Chen *et al.*, 2023; Zhong *et al.*, 2024; Cheng *et al.*, 2025]. Some studies [Zhang *et al.*, 2018; Chen *et al.*, 2023] focus on designing fusion networks on visual, textual, and user information for content popularity estimation. Recent work [Zhong *et al.*, 2024; Cheng *et al.*, 2024a; Xu *et al.*, 2025b] proposes using a multimodal retrieval-augmented framework for enhancing prediction performance on social media datasets. These methods demonstrate the benefits of multimodal fusion and retrieval augmentation, but they focus on static content and user features, overlooking cascade propagation dynamics.

6 Conclusion

In this work, we propose CasUGC, a novel framework for cascade popularity prediction that explicitly models the interplay between user-generated comment evolution and structural–temporal dynamics. To bridge the gap between heterogeneous representation spaces and mitigate alignment uncertainty, we introduce a dual-granularity alignment mechanism. Furthermore, a dynamic cascade feature generation module is designed to capture propagation uncertainty and produce temporally enhanced cascade embeddings. Extensive experiments on four real-world social media datasets demonstrate that CasUGC consistently outperforms state-of-the-art baselines, confirming the critical role of unifying comment semantics with cascade dynamics for robust forecasting.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grants No.62572119 and 62232004, Jiangsu Provincial Key Laboratory of Network and Information Security under Grants No.BM2003201, Key Laboratory of Computer Network and Information Integrationof Ministry of Education of China under Grants No.93K-9, and partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization, Collaborative Innovation Center of Wireless Communications Technology. We also thank the Big Data Computing Center of Southeast University for providing the experiment environment and computing facility.

References

[Bandari *et al.*, 2012] Roja Bandari, Sitaram Asur, and Bernardo A Huberman. The pulse of news in social media: Forecasting popularity. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, ICWSM '12, pages 26–33, 2012.

[Bao *et al.*, 2013] Peng Bao, Hua-Wei Shen, Junming Huang, and Xue-Qi Cheng. Popularity prediction in microblogging network: a case study on sina weibo. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, pages 177–178, 2013.

[Bao *et al.*, 2024] Peng Bao, Rong Yan, and Caipiao Yang. Popularity prediction via modeling temporal dependencies on dynamic evolution process. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):6828–6838, 2024.

[Cao *et al.*, 2017] Qi Cao, Huawei Shen, Keting Cen, Wentao Ouyang, and Xueqi Cheng. Deephawkes: Bridging the gap between prediction and understanding of information cascades. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, pages 1149–1158, 2017.

[Chen *et al.*, 2023] Xiaolu Chen, Weilong Chen, Chenghao Huang, Zhongjian Zhang, Lixin Duan, and Yanru Zhang. Double-fine-tuning multi-objective vision-and-language transformer for social media popularity prediction. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, pages 9462–9466, 2023.

[Cheng *et al.*, 2014] Justin Cheng, Lada A. Adamic, P. Alex Dow, Jon Kleinberg, and Jure Leskovec. Can cascades be predicted. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 925–936, 2014.

[Cheng *et al.*, 2024a] Zhangtao Cheng, Jienan Zhang, Xovee Xu, Goce Trajcevski, Ting Zhong, and Fan Zhou. Retrieval-augmented hypergraph for multimodal social media popularity prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pages 445–455, 2024.

[Cheng *et al.*, 2024b] Zhangtao Cheng, Fan Zhou, Xovee Xu, Kunpeng Zhang, Goce Trajcevski, Ting Zhong, and Philip S Yu. Information cascade popularity prediction via probabilistic diffusion. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):8541–8555, 2024.

[Cheng *et al.*, 2025] Zhangtao Cheng, Jiao Li, Jian Lang, Ting Zhong, and Fan Zhou. In-context prompt-augmented micro-video popularity prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI '25, pages 11527–11535, 2025.

[Donnat *et al.*, 2018] Claire Donnat, Marinka Zitnik, David Hallac, and Jure Leskovec. Learning structural node embeddings via diffusion wavelets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '18, pages 1320–1329, 2018.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of the Neural Information Processing Systems*, NeurIPS '20, pages 6840–6851, 2020.

[Hsu *et al.*, 2019] Chih-Chung Hsu, Li-Wei Kang, Chia-Yen Lee, Jun-Yi Lee, Zhong-Xuan Zhang, and Shao-Min Wu. Popularity prediction of social media based on multi-modal feature mining. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, pages 2687–2691, 2019.

[Jiao *et al.*, 2024] Pengfei Jiao, Hongqian Chen, Qing Bao, Wang Zhang, and Huaming Wu. Enhancing multi-scale diffusion prediction via sequential hypergraphs and adversarial learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI '24, pages 8571–8581, 2024.

[Jing *et al.*, 2025] Xin Jing, Yichen Jing, Yuhuan Lu, Bangchao Deng, Xueqin Chen, and Dingqi Yang. Casft: Future trend modeling for information popularity prediction with dynamic cues-driven diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI '25, pages 11906–11914, 2025.

[Li *et al.*, 2017] Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. Deepcas: An end-to-end predictor of information cascades. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17', pages 577–586, 2017.

[Li *et al.*, 2018] Cheng Li, Xiaoxiao Guo, and Qiaozhu Mei. Joint modeling of text and networks for cascade prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM '18, pages 640–643, 2018.

[Lu *et al.*, 2023] Xiaodong Lu, Shuo Ji, Le Yu, Leilei Sun, Bowen Du, and Tongyu Zhu. Continuous-time graph learning for cascade popularity prediction. *arXiv preprint arXiv:2306.03756*, 2023.

[Ma *et al.*, 2013] Zongyang Ma, Aixin Sun, and Gao Cong. On predicting the popularity of newly emerging hashtags in t witter. *Journal of the American Society for Information Science and Technology*, 64(7):1399–1410, 2013.

[Pinto *et al.*, 2013] Henrique Pinto, Jussara M. Almeida, and Marcos A. Gonçalves. Using early view patterns to predict the popularity of youtube videos. In *Proceedings of* *the 22nd ACM International Conference on Information and Knowledge Management*, CIKM '13, pages 365–374, 2013.

[Radford *et al.*, 2021] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning*, ICML '21, pages 8748–8763, 2021.

[Reimers and Gurevych, 2019] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[Rizoiu *et al.*, 2017] Marian-Andrei Rizoiu, Lexing Xie, Scott Sanner, Manuel Cebrian, Honglin Yu, and Pascal Van Hentenryck. Expecting to be hip: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 735–744, 2017.

[Shen *et al.*, 2014] Huawei Shen, Dashun Wang, Chaoming Song, and Albert-László Barabási. Modeling and predicting popularity dynamics via reinforced poisson processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI '14, pages 291–297, 2014.

[Suh *et al.*, 2010] Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 177–180, 2010.

[Sun *et al.*, 2023a] Xigang Sun, Jingya Zhou, Ling Liu, and Wenqi Wei. Explicit time embedding based cascade attention network for information popularity prediction. *Information Processing & Management*, 60(3):103278, 2023.

[Sun *et al.*, 2023b] Xigang Sun, Jingya Zhou, Ling Liu, and Zhen Wu. Castformer: A novel cascade transformer towards predicting information diffusion. *Information Sciences*, 648:119531, 2023.

[Szabó and Huberman, 2010] Gábor Szabó and Bernardo A Huberman. Predicting the popularity of online content. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 271–280, 2010.

[Velickovic *et al.*, 2017] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

[Wang *et al.*, 2013] Dashun Wang, Chaoming Song, and Albert-László Barabási. Quantifying long-term scientific impact. *Science*, 342(6154):127–132, 2013.

[Wu *et al.*, 2022] Zhen Wu, Jingya Zhou, Ling Liu, Chaozhuo Li, and Fei Gu. Deep popularity prediction in multi-source cascade with heri-gcn. In *2022 IEEE 38th International Conference on Data Engineering*, ICDE '22, pages 1714–1726, 2022.

[Xu *et al.*, 2021] Xovee Xu, Fan Zhou, Kunpeng Zhang, Siyuan Liu, and Goce Trajcevski. Casflow: Exploring hierarchical structures and propagation uncertainty for cascade prediction. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3484–3499, 2021.

[Xu *et al.*, 2025a] Haowei Xu, Chao Gao, Xianghua Li, and Zhen Wang. D2: Customizing two-stage graph neural networks for early rumor detection through cascade diffusion prediction. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, WSDM '25, pages 568–576, 2025.

[Xu *et al.*, 2025b] Xovee Xu, Yifan Zhang, Fan Zhou, and Jingkuan Song. Improving multimodal social media popularity prediction via selective retrieval knowledge augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI '25, pages 932–940, 2025.

[Xu *et al.*, 2025c] Yijie Xu, Bolun Zheng, Wei Zhu, Hangjia Pan, Yuchen Yao, Ning Xu, Anan Liu, Quan Zhang, and Chenggang Yan. Smtpd: A new benchmark for temporal prediction of social media popularity. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, CVPR '25, pages 18847–18857, 2025.

[Yu *et al.*, 2015] Linyun Yu, Peng Cui, Fei Wang, Chaoming Song, and Shiqiang Yang. From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics. In *Proceedings of the 2015 IEEE International Conference on Data Mining*, ICDM '15, pages 559–568, 2015.

[Zhang *et al.*, 2018] Wei Zhang, Wen Wang, Jun Wang, and Hongyuan Zha. User-guided hierarchical attention network for multi-modal social image popularity prediction. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 1277–1286, 2018.

[Zhao *et al.*, 2015] Qingyuan Zhao, Murat A. Erdogdu, Hera Y. He, Anand Rajaraman, and Jure Leskovec. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1513–1522, 2015.

[Zhong *et al.*, 2024] Ting Zhong, Jian Lang, Yifan Zhang, Zhangtao Cheng, Kunpeng Zhang, and Fan Zhou. Predicting micro-video popularity via multi-modal retrieval augmentation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '24, pages 2579–2583, 2024.

[Zhou *et al.*, 2021] Fan Zhou, Xovee Xu, Goce Trajcevski, and Kunpeng Zhang. A survey of information cascade analysis: Models, predictions, and recent advances. *ACM Computing Surveys*, 54(2):1–36, 2021.

[Zhou *et al.*, 2024] Mingyang Zhou, Yanjie Lin, Gang Liu, Zuwen Li, Hao Liao, and Rui Mao. Modeling personalized retweeting behaviors for multi-stage cascade popularity prediction. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, IJCAI '24, pages 2598–2606, 2024.