

Evolving Interest for Information Diffusion Prediction on Social Network

Yuyang Liu^{*†}, Junruo Gao^{*†}, Zefang Zhao^{*†}, Haibo Wu^{*}, Zhaojuan Yue^{*} and Jun Li^{*‡}

^{*}Computer Network Information Center, Chinese Academy of Sciences, Beijing, China

[†]University of Chinese Academy of Sciences, Beijing, China

{liuyuyang, gaojunruo, zhaozefang, wuhaibo, yuezhaojuan, lijun}@cnic.cn, [‡]Corresponding Author

Abstract—Even though conventional methods have contributed a lot to predicting information diffusion utilizing an end-to-end framework, they omit to consider the reason why each participant is involved in the cascade. Inspired by the ubiquitous pattern that users are inclined to take part in the discussion with attractive content and like-minded participants, we propose a temporal evolving interest-driven cascade prediction framework, named TEIC. The proposed TEIC is capable of automatically capturing the interest-driven forwarding behavior of individuals from the micro perspective and assembling them for macro-level cascade size prediction. Taking historical social actions as the input, the temporal evolving interest encoder is designed to characterize individual social preferences that change dynamically over time. Furthermore, we adopt a cascade aggregator to integrate microscopic interest-driven social actions into macroscopic cascade representations for predicting the incremental diffusion scale. We compare the TEIC with the multiple baselines, including hand-crafting feature regression, generative methods and deep learning-based models. Extensive experiments on two real-world datasets demonstrate that the proposed model significantly outperforms state-of-the-art approaches.

Keywords—social network, information diffusion, cascade prediction, neural network, evolving interest

I. INTRODUCTION

The global penetration of online social service, e.g., Facebook, LinkedIn and Instagram, makes information diffusion increasingly unimpeded. Various types of content can be generated or reshared according to the inclination of users; this is a micro-explanation of the information cascade. From a macroscopic perspective, the term “information cascade” mainly refers to the process in which attractive content is hierarchically forwarded by individuals after it is released. An increasing number of research has been conducted on modeling information diffusion due to the immense potentiality in varied scenarios, including strategic marketing [1], influence maximization [2] and rumor detection [3], [4]. Information cascade prediction has clearly become a prevailing yet complicated method of providing instructive insights into the development of strategies in advance; thus, methodologies to model and understand the inherent dynamics of information diffusion are urgently needed.

Initially, taking into account the expert experience of network evolution and information dissemination, whether an information cascade can be predicted and what role various features play in the propagation process are verified [5]. As a

further step to explain the underlying mechanism of cascade formation, the differential equations are introduced to model the state transmission of individuals [6], [7] or communities [8], [9]. Currently, enlightened by the surpassing triumph of neural networks in various applications, graph structure and its dynamic evolution are fully represented by end-to-end frameworks [10], [11], especially recurrent and graph neural networks. Even though these advanced frameworks provide an intuitive grasp of the complex networks, they are mainly limited to the single paradigm that focuses on recognizing the common propagation patterns from temporal and topological data and ignore the essence of articulating social reality. People and their behavior, as the determinants of information propagation, deserve more exploration. For example, when users browse a large number of messages on social networks, which messages will attract users’ attention and be forwarded sequentially? Users are more willing to take part in discussions they are interested in, where like-minded early participants will further incentivize the user to participate. This concept is further crystallized as a quantitative issue: can we accurately predict the scale of information cascade with users’ interest-driven forwarding behavior as the microscopic impetus?

To tackle the above problem, we propose a temporal evolving interest-driven cascade prediction framework, named TEIC. By architecting the social action representing, temporal evolving interest encoder and cascade aggregation into a unified framework, we expect the proposed model to automatically discover the latent connection between interest-driven social behavior and information diffusion. In summary, the main contributions of our work are as follows:

- We explore how individual interests evolve over time and design an interest encoder, which extracts interest and semantic information from the users’ historical social actions by a temporal evolving attention mechanism.
- We introduce a sequential aggregation strategy that integrates all micro-level forwarding actions into a macro-level cascade representation for scale prediction. Additionally, the sampled representative non-participants are taken into account as auxiliary information.
- We conduct extensive evaluations on two publicly available benchmark datasets—Sina Weibo and OAG 2.0, demonstrating that the proposed TEIC outperforms the state-of-the-art baselines.

Organization The rest of the paper is organized as follows. A general discussion of information cascade prediction and user interest representation are given in Section II; Section ?? defines the information cascade prediction problem. Following, Section IV describes the TEIC framework. The extensive results are discussed in Section V. Finally, Section VI concludes the proposed model.

II. RELATED WORK

Existing studies that have been proposed to cope with the cascade prediction issue can be grouped into three categories, including:

Feature-based approaches mainly focus on how to extract and coalesce complicated features, including temporal features [12], topological features [13]–[15], user features [16]–[19], etc. It is also important to design sophisticated models that can effectively combine these hand-crafted features and specific tasks. In order to explore how to control the misinformation dissemination on temporal social networks when the cost of intervention is limited, [20] designed a sophisticated centrality metric, called the backbone link property. From the perspective of diffusion and social influence theory, the multi-dimensional features of all information dissemination actors are incorporated into cascade modeling to predict the trustworthiness of a given online content [18]. These feature-based methods are transferred to other domains with difficulty since such methods are subject to complex manual feature engineering.

Diffusion-based approaches typically represent the relationship between pair-wise individuals or communities as the vehicle for message dissemination. Generative methods [6], [7] focus on regarding information diffusion as typical stochastic processes, such as Poisson process and Hawkes process. These stochastic processes generally model each user as a time-dependent intensity function, whose instantaneous intensity value changes correspondingly as different users participate in the propagation. [21] attempted to explore the reasons for each forwarding behavior from a sociological point of view. They divided the forwarding motivations into content-driven forwarding and conformity-driven forwarding, which are incorporated into the design of the Hawkes intensity function to achieve link prediction. The intensity functions and corresponding covariates in these methods provide an intuitive explanation for understanding the underlying diffusion dynamics. However, their predictive power is still far from satisfactory.

Deep learning-based approaches, especially recurrent neural network (RNN)-based [10], [22], [23] and graph neural network (GNN)-based [24] models, can automatically capture sequential and graphical characteristics. DeepCas [23], as the first end-to-end framework that introduces deep learning technology into information propagation prediction, separates a complete diffusion graph into multiple ordered node lists by random walk and assembles these sequences using a geometric attention mechanism. Subsequently, the generative Hawkes process was reimplemented using a deep sequential learning approach [22]. This combination balances the interpretability

and prediction accuracy. Inspired by the excellent performance of graph convolutions, both directed and undirected graph convolutions are introduced into the diffusion structure modeling [11], [24]. These deep learning-based models achieve disentanglement of hand-crafted feature screening while achieving better prediction performance.

III. PRELIMINARIES

For a specific message s , the corresponding cascade $\mathcal{C}(s, t) = \{(u_1, t_1), (u_2, t_2), \dots, (u_N, t_N) | t_N < t\}$ is a sequence of forwarding tuples (u_n, t_n) , where t_n is the timestamp when individual u_n posts or forwards a message. Intuitively, u_1 is the original publisher of message s . The cascade size is the number of times it is forwarded up to timestamp t , denoted as $|\mathcal{C}(s, t)|$. Then, the cascade prediction problem to be solved is defined as follows: given the early stage of a cascade $\mathcal{C}(s, T)$ that is observed during window $[0, T)$, the increment size $\Delta\Phi = |\mathcal{C}(s, \infty)| - |\mathcal{C}(s, T)|$ is predicted, where $|\mathcal{C}(s, \infty)|$ can represent the final cascade size.

In order to tackle the above prediction problem based on auxiliary interest information, we additionally introduce some notations related to individual historical social behavior. As for each participant u_n in cascade $\mathcal{C}(s, t)$, its historical action sequence is defined as $\{(m_n^1, \tau_n^1), (m_n^2, \tau_n^2), \dots, (m_n^K, \tau_n^K) | \tau_n^K < t_n\}$, where τ_n^k is the timestamp when individual $u_n \in \mathcal{C}(s, t)$ posts or forwards the historical message m_n^k . The historical action sequence represents other messages that user u_n has forwarded before participating in cascade $\mathcal{C}(s, t)$.

IV. PROPOSED METHOD

In this section, we will introduce the proposed TEIC framework in detail, which contains two parts: (1) evolving interest encoder, which extracts interest and semantic information from the users' historical actions by temporal evolving attention mechanism; and (2) cascade aggregation, which assembles interest-driven social actions while taking into account non-participants of cascade.

A. Evolving Interest Encoder

User interests affect information diffusion in a manner resembling catalysts, which incline a message to spread in an interest-based community. In this subsection, we set out to extract evolving interests from users' historical social actions.

Given a participant u_n in cascade $\mathcal{C}(s, t)$, whose historical actions consists of a sequence of forwarding behavior $\{(m_n^1, \tau_n^1), (m_n^2, \tau_n^2), \dots, (m_n^K, \tau_n^K)\}$, we first convert each message m_n^k into its low-dimensional vector using pre-trained language model and get a vector e_n^k . The latent message vector is further compressed by average pooling:

$$\mathbf{E}_n^k(\mathbf{m}) = \text{Avg}(e_n^k). \quad (1)$$

The timestamp τ_n^k is embedd into latent vectors $\mathbf{E}_n^k(\tau)$ based on the time2vec. Thus, the tuple of historical social action can be rewritten as $(\mathbf{E}_n^k(\mathbf{m}), \mathbf{E}_n^k(\tau))$. We also explore the effect of relative position of historical actions. The index

embedding $\mathbf{P}^k$ is constructed by performing periodic trigonometric transformation of the positional variables k :

$$\mathbf{P}^k[i] = \begin{cases} \sin(k/10000^{2i/d}) & \text{if } i \text{ is even,} \\ \cos(k/10000^{2i/d}) & \text{if } i \text{ is odd,} \end{cases}$$

where d is the number of latent dimension and $i \leq d/2$ is the specific dimensional index.

Then, each historical action of participant u_n is generated as:

$$\mathbf{E}_n^k = \mathbf{E}_n^k(\mathbf{m}) \oplus \mathbf{E}_n^k(\tau) \oplus \mathbf{P}^k, \quad (2)$$

where $\oplus$ represents the element-wise summation.

We feed all action embeddings to the multi-head temporal evolving attention layer, which is designed to capture the temporal evolving interest from multiple perspectives. Concretely, we first calculate the contribution of different actions:

$$\alpha_{n,i}^k = \mathbf{E}_n^k \mathbf{W}_i^Q (\mathbf{E}_n^k \mathbf{W}_i^O)^T / \sqrt{d}, \quad (3)$$

where $\mathbf{W}_i^Q$ and $\mathbf{W}_i^O$ are the projection parameters of the i -th attention head respectively.

The influence of interests at different historical stages on users' current forwarding action is inconsistent. Generally speaking, the closer the timestamp τ_n^k of message m_n^k is to the current forwarding timestamp t_n , the greater the impact. Thus, we design an power-law evolving process for all historical actions as:

$$\hat{\alpha}_{n,i}^k = \frac{(\tau_n^k)^\theta \alpha_{n,i}^k}{\sum_{j=1}^K (\tau_n^j)^\theta \alpha_{n,i}^j}. \quad (4)$$

Thus, all historical actions is aggregated by weighted summation:

$$\hat{\mathbf{E}}_n = \sum_{k=1}^K (\hat{\alpha}_{n,i}^k \mathbf{E}_n^k \mathbf{W}_i^V), \quad (5)$$

We concatenate multiple head representation related to one specific user and adopt a linear projection:

$$\mathbf{U}_n = \left(\mathbf{W}_u \left\| \hat{\mathbf{E}}_n \right\|_{i=1}^I \right) + \mathbf{b}_u, \quad (6)$$

where $\|$ is the concatenation operation, and $\mathbf{W}_u$ and $\mathbf{b}_u$ are weight matrices and bias, respectively.

B. Cascade Aggregation

After getting the latent interest embeddings of users, the interest-driven forwarding actions will be aggregated into macro cascade representation. Inspired by previous information diffusion model [25], we not only consider participants within the observation window, but also take into account some typical non-participants that serve as negative samples.

In addition to the normalized low-dimensional embedding that comes from the upstream layer, the input $\hat{\mathbf{U}}_n$ for u_n covers one auxiliary variable,

$$\hat{\mathbf{U}}_n = \mathbf{U}_n \parallel \gamma_n, \quad (7)$$

where γ_n is a binary variable that indicates whether the user is a participant or non-participant.

All participants are fed into the gate recurrent units in chronological order of the forwarding action. Specific to the n -th forwarder, we introduce how the gate recurrent unit calculates the interest-driven forwarding behavior. The previous hidden state $\mathbf{h}_{n-1}$ records the overall picture of information cascade before timestamp t_n . The current input $\hat{\mathbf{U}}_n$ is associated with user u_i and its evolving interest at timestamp t_n .

The neural unit models the participation process of user u_n and updates the cumulative representation of the corresponding cascade by means of a gating mechanism. First, taking into account the relationship between u_n 's interest and the overall picture of the previous cascade, user u_n 's interest-driven forwarding behavior is computed as follows:

$$\tilde{\mathbf{h}}_n = \tanh(\mathbf{W}_h \hat{\mathbf{U}}_n + \mathbf{r}_n \odot (\mathbf{V}_h \mathbf{h}_{n-1}) + \mathbf{b}_h), \quad (8)$$

where $\odot$ represents an element-wise product and the reset gate $\mathbf{r}_n$ is computed by

$$\mathbf{r}_n = \sigma(\mathbf{W}_r \hat{\mathbf{U}}_n + \mathbf{V}_r \mathbf{h}_{n-1} + \mathbf{b}_r). \quad (9)$$

After user u_n is involved in the cascade, the overall picture of the previous cascade is also updated:

$$\mathbf{h}_n = \mathbf{z}_n \odot \mathbf{h}_{n-1} + (1 - \mathbf{z}_n) \odot \tilde{\mathbf{h}}_n, \quad (10)$$

where the update gate $\mathbf{z}_n$ is computed by

$$\mathbf{z}_n = \sigma(\mathbf{W}_z \hat{\mathbf{U}}_n + \mathbf{V}_z \mathbf{h}_{n-1} + \mathbf{b}_z). \quad (11)$$

For each timestamp t_n , the hidden state $\mathbf{h}_n$ is updated based on user u_n 's participation, which represents the latest development of cascade. In contrast to the gate units corresponding to participants, the remaining units model the non-participant terms. When observation time T is fixed, we empirically sample the most representative non-participants from the users who keep directed link with source user u_1 but do not participant in the cascade (s, T) before T . The sampled non-participants are fed into gate recurrent units in descending order of the number of interactions with the root user.

It should be noted that the excitation of each observed user, including participants and non-participants, will influence differently the incremental scale of the message. Thus, we attempt to use the cohesion degree between each user and the source message s to measure their contributions. More formally, the coefficient is computed as follows:

$$\beta_n = \frac{\mathbf{h}_n \mathbf{W}_s \mathbf{E}(s)}{\sum_{j=1}^N \mathbf{h}_j \mathbf{W}_s \mathbf{E}(s)}, \quad (12)$$

where $\mathbf{E}(s)$ is the text embedding of the source message s .

TABLE I
STATISTICS OF THE DATASETS.

	Sina Weibo	OAG 2.0
Users	1,776,950	1,047,324
Messages	300,000	89,572
$Avg. \Delta\Phi$	42.23	21.76

Once the normalized coefficient is obtained, we assemble the snapshots at each timestamp by weighted sum to characterize the diffusion dynamics over time:

$$C_{s,T} = \sum_{n=1}^N \beta_n h_n. \quad (13)$$

C. Output Layer

The last component of TEIC is a feed forward network with one final output unit. Given the representation $C_{s,T}$, we predict the incremental size of the cascade, which can be achieved by minimizing the following loss function:

$$\mathcal{L} = \frac{1}{|S|} \sum_{s \in S} (\log \text{FFN}(C_{s,T}) - \log \Delta\Phi)^2, \quad (14)$$

where S is entire cascade set and $\Delta\Phi$ is the ground truth.

V. EXPERIMENT

In this section, we describe the experimental setup and provide a detail discussion on performance comparison.

A. Datasets

We mainly consider the social network and academic citation network—Sina Weibo [26] and OAG 2.0 [27], [28], both of which contain a large amount of information dissemination in the dynamic evolution process. Table I summarizes the characteristics of two datasets.

Sina Weibo [26] is the most popular social tool in China. This platform has about 573 million monthly active users. A huge amount of user-generated content is browsed and forwarded every day.

Open academic graph 2.0 (OAG 2.0) extracts 64,639,608 common academic articles in Microsoft Academic Graph [27] and Aminer [28], and marks the citation relationship between the articles similar to the forwarding behavior on social platforms. In the experiments, we only use abstracts for calculation.

B. Evaluation Metric

Accuracy Measure: Following the existing work [11], we utilize the mean square logarithmic error (MSLE) to evaluate the accuracy of the predicted result.

Hyper-parameter Analysis: We explore how the prediction performance changes with respect to the continuous adjustment of hyper-parameters.

TABLE II
LIST OF FEATURES USED IN BASELINES.

Type	Description
Temporal Features	The average time interval between two retweets.
	The cumulative scale every fixed interval.
	The incremental scale every fixed interval.
Structural Features	The number of leaf nodes.
	The average degree.
	The average length of sub-path.
	The max length of sub-path.
User & Content Features	The user identity.
	The number of fans and friends of each user.
	The number of historical posts.

C. Comparison Methods

We consider five different benchmarks to compare TEIC against. Selected baselines cover feature regression, generative point process and neural network-based models.

Feature-based LR: Linear regression (LR) is adopted to compute the relationship between the handcrafted features and cascade size, as shown in Table II. Moreover, we append a weighted L2 regularization term to penalize the model complexity.

NEWER [29]: NEWER ingeniously integrates survival analysis into behavioral dynamics and aggregates the one-hop subcascades for scale prediction.

DeepCas [23]: DeepCas is a neural framework that relies on the structures of cascading graphs and the identities of participating users.

DeepHawkes [22]: DeepHawkes not only retains the high accuracy from deep learning but also inherits the interpretability from the generative model.

CasCN [11]: CasCN is a convolution method suitable for directed graphs. The graph embeddings at different timestamps are fed into a recurrent neural network to learn previously missed temporal dynamics.

D. Hyper-parameter Setting

For our model TEIC, we first select the recent 20 historical records for each user and only intercept the front part with fixed length. The BERT [30] is adopted for message embedding. In the evolving interesting layer, the encoder converts historical actions into a 64-dimensional interest embedding. For the aggregation part, each gating unit has 32 units. All the parameters are initialized with He initialization [31] and trained using the Adam [32] optimizer. For each dataset, we randomly select 80% of historical information cascades to constitute the training set, 10% as the validation set, and treat the remaining cascades as the test set. The batch size for each iteration is 32 and the training process will stop when the loss of validation set does not decline for 10 consecutive iterations. For the other deep learning-based models, we follow the settings in DeepCas [23], where the embedding size of the users equals 50 and the hidden layer of each GRU has 32 units.

TABLE III
OVERALL PREDICTION PERFORMANCE OF THE INCREMENT SIZE AMONG DIFFERENT METHODS IN TERMS OF THE MSLE.

Datasets	Sina Weibo			OAG 2.0		
T	1 hour	2 hours	3 hours	1 year	3 years	5 years
Feature-based LR	3.693	3.568	3.584	2.739	2.668	2.523
NEWER	3.529	3.341	3.362	2.517	2.370	2.311
DeepCas	3.301	3.147	2.960	2.164	1.969	1.814
DeepHawkes	2.548	2.343	2.291	2.159	1.853	1.829
CasCN	2.417	2.280	2.183	1.884	1.762	1.709
TEIC	2.362	2.217	2.066	1.823	1.730	1.639

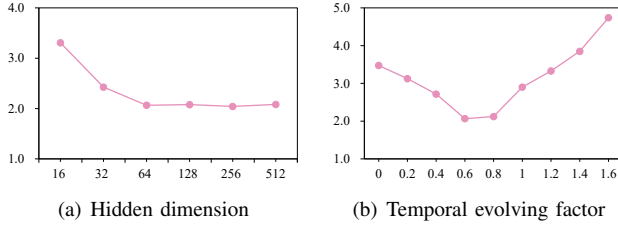


Fig. 1. Hyper-parameters Analysis on the Sina Weibo Dataset.

E. Performance Comparison

As shown in Table III, TEIC outperforms other state-of-the-art baselines, illustrating the effectiveness of our proposed framework.

The prediction precision of NEWER is slightly improved relative to feature-based linear regression. Both methods can be regarded as techniques for generating and utilizing hand-crafted features based on historical experience and observations. Heavily relying on the priori knowledge potentially imposes strong constraints on the two models. Deep learning-based methods are superior to the above feature-based approach and generative approach to varying degrees. As the first proposed end-to-end learning strategy, DeepCas eliminates the dependence on hand-crafted features, but it is much worse than others because it ignores temporal features, which represent the most predictive information. DeepHawkes and CasCN are comparable in performance. Although DeepHawkes chooses more adequate features, it makes a trade-off between prediction performance and interpretability rather than purely pursuing accuracy. CasCN could fully learn topological information with the assistance of the multi-directional GCN. Nevertheless, CasCN lacks a fine-grained characterization of the user's forwarding behavior because it analyzes the forwarding graph as a whole at each timestamp. Finally, our proposed TEIC achieves excellent prediction results, with $MSLE = 2.066$ on Sina Weibo and $MSLE = 1.639$ on OAG 2.0 when the observation window equals 3 hour and 7 years respectively.

F. Parameter Analysis

We evaluate how TEIC performs against different hyper-parameters.

Hidden dimension. We vary the size of the hidden dimensions in evolving interest encoder from 16 to 32, 64, 128, 256, and 512. As shown in Figure 1(a), the proposed

TABLE IV
PERFORMANCE COMPARISON BETWEEN TEIC AND ITS VARIANTS.

Datasets	Sina Weibo			OAG 2.0		
Δt	1 hour	2 hours	3 hours	3 years	5 years	7 years
TEIC/MA	3.798	3.659	3.529	2.938	2.886	2.781
TEIC/TE	3.347	3.149	2.876	2.724	2.545	2.319
TEIC/NP	3.134	2.786	2.507	2.368	2.109	1.801
TEIC	2.362	2.217	2.066	1.823	1.730	1.639

framework achieves best performance when the dimensionality of embedding U_n equals 64.

Temporal evolving factor. The evolving factor θ in Equation 4 determines the influence of dynamic interests evolving over time on users' latest social intentions. We vary the evolving factor from 0 to 2. The results are shown in Figure 1(b). There is a peak when the factor is 0.6.

G. Ablation Study

To substantiate the effectiveness of the major components in the IntCas model, we present three variants, denoted as **TEIC/MA**, **TEIC/TE** and **TEIC/NP**. **TEIC/MA** removes multiple historical actions from TEIC and uses latest one action to represent user interests. **TEIC/TE** removes the temporal evolving mechanism form TEIC. In **TEIC/NP**, the non-participant terms are omitted. Table IV outlines the performance comparison among TEIC and its variants. TEIC/MA and TEIC/TE both simplify the modeling of user interests, whose prediction accuracy is far from satisfactory. For TEIC-NP, the omission of non-participants results in a minor impact on performance that corroborates the assumptions of previous works [7].

VI. CONCLUSION

Motivated by the need of effectively predicting cascading process, we study information propagation dynamics. We formulate this problem from the perspective of user interests and propose a temporal evolving interest-driven cascade prediction framework, by incorporating user interest representation and cascade aggregation techniques that do not heavily depend on feature engineering. We evaluate TEIC on both social and academic network datasets. Compared to the state of the art information diffusion prediction algorithms that pay more attention to temporal and structural features, TEIC achieves significant and robust error reduction. This work explores the potential of user interest in the cascade size prediction problem.

VII. ACKNOWLEDGMENTS

This work was funded by the National Key Research and Development Project of China (No. 2019YFB1405801) and the International Partnership Program of the Chinese Academy of Sciences (No. 241711KYSB20180002).

REFERENCES

- [1] S. Tu and S. Neumann, "A viral marketing-based model for opinion dynamics in online social networks," in *Proceedings of the ACM Web Conference 2022*, 2022, p. 1570–1578.
- [2] S. S. Singh, D. Srivastva, M. Verma, and J. Singh, "Influence maximization frameworks, performance, challenges and directions on social network: A theoretical study," *Journal of King Saud University - Computer and Information Sciences*, 2021.
- [3] N. Rani, P. Das, and A. K. Bhardwaj, "Rumor, misinformation among web: A contemporary review of rumor detection techniques during different web waves," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 1, p. e6479, 2022.
- [4] W. Zhang, T. Zhong, C. Li, K. Zhang, and F. Zhou, "Causalrd: A causal view of rumor detection via eliminating popularity and conformity biases," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, 2022, pp. 1369–1378.
- [5] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec, "Can cascades be predicted?" in *Proceedings of the 23rd international conference on World wide web*, 2014, pp. 925–936.
- [6] Z. Pan, Z. Huang, D. Lian, and E. Chen, "A variational point process model for social event sequences," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 173–180, Apr. 2020.
- [7] B. Zhou, S. Pei, L. Muchnik, X. Meng, X. Xu, A. Sela, S. Havlin, and H. E. Stanley, "Realistic modelling of information spread using peer-to-peer diffusion patterns," *Nature Human Behaviour*, vol. 4, no. 11, pp. 1198–1207, 2020.
- [8] N. Sivaraman, "Exo-sir: An epidemiological model to quantify the exogenous information diffusion and its application to detect events," in *13th ACM Web Science Conference 2021*, 2021, p. 145–146.
- [9] T. Chen, J. Shi, J. Yang, G. Cong, and G. Li, "Modeling public opinion polarization in group behavior by integrating sirs-based information diffusion process," *Complexity*, vol. 2020, pp. 4 791 527:1–4 791 527:20, 2020.
- [10] X. Tang, D. Liao, W. Huang, J. Xu, L. Zhu, and M. Shen, "Fully exploiting cascade graphs for real-time forwarding prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 1, pp. 582–590, May 2021.
- [11] X. Chen, F. Zhou, K. Zhang, G. Trajcevski, T. Zhong, and F. Zhang, "Information diffusion prediction via recurrent cascades convolution," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 2019, pp. 770–781.
- [12] A. K. Bhowmick, "Temporal pattern of retweet(s) help to maximize information diffusion in twitter," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, 2020, p. 913–914.
- [13] C. Liu, N. Zhou, X.-X. Zhan, G.-Q. Sun, and Z.-K. Zhang, "Markov-based solution for information diffusion on adaptive social networks," *Applied Mathematics and Computation*, vol. 380, p. 125286, 2020.
- [14] J. Chen, X. Xu, L. Chen, Z. Ruan, Z. Ming, and Y. Liu, "Ctl-diff: Control information diffusion in social network by structure optimization," *IEEE Transactions on Computational Social Systems*, pp. 1–15, 2022.
- [15] G. Fu, F. Chen, J. Liu, and J. Han, "Analysis of competitive information diffusion in a group-based population over social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 525, pp. 409–419, 2019.
- [16] Y. Wang, J. Wang, H. Wang, R. Zhang, and M. Li, "Users' mobility enhances information diffusion in online social networks," *Information Sciences*, vol. 546, pp. 329–348, 2021.
- [17] L. Sun, Y. Rao, X. Zhang, Y. Lan, and S. Yu, "Ms-hgat: Memory-enhanced sequential hypergraph attention network for information diffusion prediction," *Proceedings of the AAAI Conference on Artificial Intelligence*, May 2022.
- [18] X. Chen, F. Zhou, F. Zhang, and M. Bonsangue, "Catch me if you can: A participant-level rumor detection framework via fine-grained user representation learning," *Information Processing & Management*, vol. 58, no. 5, p. 102678, 2021.
- [19] R. Wang, Z. Huang, S. Liu, H. Shao, D. Liu, J. Li, T. Wang, D. Sun, S. Yao, and T. Abdelzaher, "Dydiff-vae: A dynamic variational framework for information diffusion prediction," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, p. 163–172.
- [20] X.-X. Zhan, A. Hanjalic, and H. Wang, "Suppressing information diffusion via link blocking in temporal networks," in *Complex Networks and Their Applications VIII*, 2020, pp. 448–458.
- [21] H. Li, H. Li, and S. S. Bhowmick, "Chassis: Conformity meets online information diffusion," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 2020, p. 1829–1840.
- [22] Q. Cao, H. Shen, K. Cen, W. Ouyang, and X. 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*, 2017, p. 1149–1158.
- [23] C. Li, J. Ma, X. Guo, and Q. Mei, "Deepcas: An end-to-end predictor of information cascades," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, p. 577–586.
- [24] Q. Cao, H. Shen, J. Gao, B. Wei, and X. Cheng, "Popularity prediction on social platforms with coupled graph neural networks," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, 2020, p. 70–78.
- [25] M. Gomez-Rodriguez, J. Leskovec, and B. Schölkopf, "Modeling information propagation with survival theory," in *International Conference on Machine Learning*, 2013, pp. 666–674.
- [26] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li, "Social influence locality for modeling retweeting behaviors," in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, 2013, p. 2761–2767.
- [27] A. Sinha, Z. Shen, Y. Song, H. Ma, D. Eide, B.-J. P. Hsu, and K. Wang, "An overview of microsoft academic service (mas) and applications," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, p. 243–246.
- [28] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: Extraction and mining of academic social networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, p. 990–998.
- [29] L. Yu, P. Cui, F. Wang, C. Song, and S. Yang, "From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics," in *2015 IEEE International Conference on Data Mining*, 2015, pp. 559–568.
- [30] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, p. 1026–1034.
- [32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.



Yuyang Liu received his Bachelor's degree from Northwestern Polytechnical University. Now he is a Ph.D. student in the University of Chinese Academy of Sciences. His research focuses on social network analysis and data mining.



Junruo Gao received her Bachelor's degree from Sichuan University. Now she is a Ph.D. student in University of Chinese Academy of Sciences. Her current research interests include artificial intelligence and recommendation system.



Zefang Zhao received his Bachelor's degree from Taiyuan University of Technology. Now he is a Ph.D. student in the University of Chinese Academy of Sciences. His research focuses on deep learning, natural language process and sentiment analysis.



Haibo Wu is an associate professor at the Computer Network Information Center of Chinese Academy of Sciences. Her research interests include network management and information diffusion modeling.



Zhaojuan Yue is a senior engineer at the Computer Network Information Center of Chinese Academy of Sciences. Her research interests include network management and information diffusion modeling.



Jun Li is a research fellow and doctoral supervisor at the Computer Network Information Center of Chinese Academy of Sciences, specially appointed researcher of Chinese Academy of Sciences. His main research interests are artificial intelligence and big data technical applications and future Internet architecture.