Public Opinion Field Effect and Hawkes Process Join Hands for Information Popularity Prediction

Junliang Li[1], Yajun Yang[1*], Yujia Zhang[2], Qinghua Hu[1], Alan Zhao[2], Hong Gao[3]

[1]College of Intelligence and Computing, Tianjin University, China
[2]AI Technology Center of OVB, Tencent, China
[3]School of Computer Science and Technology, Zhejiang Normal University, China
{lijunliang, yjyang, huqinghua}@tju.edu.cn, {yujiazhang, alantzhao}@tencent.com, honggao@zjnu.edu.cn

Abstract

Information popularity prediction, aiming to predict the growth of user participation in a trending topic diffusion, is a fundamental task in social networks. Existing methods often treat information diffusion as a single independent process, ignoring the "public opinion field effect" where multiple trending topics coexist and compete for user attention simultaneously. Inspired by Hawkes theory, we propose a novel Hawkes-process-based learning model for information popularity prediction, which takes into account both the temporal correlation among users' propagation behaviors in several topics diffusion and public opinion field effect in social networks. We first propose an improved neural Hawkes process to capture comprehensive propagation law from multiple dimensions and then propose a novel public opinion field paradigm based on the improved Hawkes process and cascade structure. We design a novel learning framework incorporating the public opinion field paradigm to extract high-quality representations for information popularity prediction. Extensive experiments on four real-world datasets validate that our model significantly outperforms the state-of-the-art competitors.

Code and Appendix — https://github.com/ki-ljl/POFHP

Introduction

Information popularity prediction is a fundamental task in social networks, which can assist in understanding the evolution of trending topics and has many important applications such as public opinion analysis, fake news controlling (Masud et al. 2021; Wang et al. 2023), and online social marketing (Aravamudan, Zhang, and Anagnostopoulos 2023).

Existing methods for popularity prediction fall into two categories: probabilistic generative and cascade learning-based. Probabilistic generative method treats information diffusion as a sequence of events (Zhao et al. 2015) and design various diffusion models to simulate the occurrence of events based on time-dependent probabilistic process, e.g., Poisson process (Daley, Vere-Jones et al. 2003) and Hawkes process (Hawkes 1971). These methods well utilize the laws of propagation in physical world for prediction and then retains good interpretability. However, these methods simplify the propagation process and neglect the influence of various factors on the topic propagation in the public opinion field. Cascade learning-based method is proposed to learn data-driven diffusion models from historical diffusion cascades. These models use recurrent neural network (RNN) (Li et al. 2017; Wang et al. 2017) and graph neural network (GNN) (Welling and Kipf 2016; Velickovic et al. 2017) respectively to capture the temporal correlation among user behaviors in cascade and the structural features of social networks. Some works combine RNN and GNN to capture cascade contexts and social networks simultaneously, e.g., CasCN (Chen et al. 2019) and DyHGCN (Yuan et al. 2021).

Both of the above two kinds of methods regard information diffusion as a single independent process. However, in real world, several trending topics always coexist in the public opinion space and compete for user attention at the same time, which deeply affects information diffusion. Recently, POFD (Li et al. 2023) is proposed to incorporate "*public opinion field effect*" for enhancing representation learning. Each topic forms a public opinion field and these fields compete for user attention in information diffusion. By modeling the competition mechanism between different public opinion fields, POFD can obtain better user representation for information diffusion prediction. Unfortunately, POFD only utilizes the network structure for representation learning but does not investigate how to use information cascade for prediction.

We propose a novel Hawkes-process-based learning model that integrates probabilistic generative and cascade-learning methods while incorporating public opinion field effects for information popularity prediction. Unlike traditional Hawkes process, which focus solely on temporal correlations and past events, our model accounts for additional factors like public opinion field effects. Inspired by the Hawkes process, we define the public opinion attraction based on the conditional intensity function and design a new public opinion field paradigm. We propose a novel neural Hawkes process to capture implicit temporal correlation among user retweeting behaviors in three dimensions: historical retweeting sequences, within a single topic, and across multiple topics. Then, a representation learning framework incorporating public opinion field effect and Hawkes process can be designed to extract high-quality node representations for information popularity prediction. The main contributions of this paper are summarized as follows.

[*]Corresponding author.

(1) We propose an improved neural Hawkes process to learn the comprehensive propagation law in information diffusion from multiple dimensions.

(2) We propose a novel public opinion field paradigm combining our improved Hawkes process and information cascade structure. We further design a representation learning framework based on this paradigm for high-quality information popularity prediction.

(3) We conduct extensive experiments on four real-world datasets. The experimental results validate the superiority of our method.

Problem Definition

This section gives a formal definition of the information popularity prediction problem.

Definition 1 (Information Cascade)**.** *Given a social network $G = (U, E)$ and a topic set $M = \{m_k | i = 1, 2, \cdots\}$ on G, U is the set of users and $E \subseteq U \times U$ represents friendship. The i-th cascade sequence of a topic m_k before time point t_o is defined as a sequence of tuples $C_{k,i}(t_o) = \{(u_j, t_j) | u_j \in U \wedge t_j < t_o\}$, where the tuple (u_j, t_j) indicates that user u_j is participated in the propagation of topic m_k at time t_j. Then, the information cascade of m_k is defined as the union of all cascade sequences, i.e., $C_k(t_o) = \bigcup_{i=1}^{R_k} C_{k,i}(t_o)$, where R_k is the number of cascade sequences of m_k.*

Example 1. *Fig. 1(a) illustrates the four cascade sequences of topic m_k: $m_k \to u_1 \to u_3$, $m_k \to u_2 \to u_5$, $m_k \to u_2 \to u_6$ and $m_k \to u_4$. The four cascade sequences consist of the information cascade $C_k(t_6)$.*

Based on information cascade, we give the definition of heterogeneous cascade graph.

Definition 2 (Heterogeneous Cascade Graph, HCG)**.** *Let heterogeneous cascade graph $G_t = (\mathcal{V}, \mathcal{E}_t) = (U \cup M, \mathcal{E}_t)$ denote the diffusion process formed by all topics up to time t, where U and M denote the set of users and topics, respectively. $\mathcal{E}_t = E \cup D_t \cup I_t$ is the set of edges, where $E \subseteq U \times U$ denotes friendship between users. $(u_i \to u_j)$ in $D_t \subseteq U \times U$ denotes the set of propagation relations, and $u_i \to m_j$ in $I_t \subseteq U \times M$ denotes that the user u_i participates in the retweeting of m_j, which is called an interest relation.*

Example 2. *Fig. 1(b) gives an example of a HCG with two kinds of nodes (users and topics) and three kinds of edges (friendship between users, propagation between users, and user interest in topics).*

Finally, based on heterogeneous cascade graph, we give a definition of information popularity prediction.

Definition 3 (Information Popularity Prediction)**.** *Given a heterogeneous cascade graph G_{t_o}, the information popularity prediction aims to predict the incremental popularity $\Delta R_{C_k} = |C_k(t_p)| - |C_k(t_o)|$ of any cascade C_k from the observation time t_o to the prediction time t_p, where $t_o \ll t_p$.*

Formalization of Public Opinion Field Effect

In order to generalize public opinion field paradigm (Li et al. 2023) to the field of information popularity prediction, in this section we give the public opinion field paradigm incorporating the Hawkes process based on HCG.

Hawkes Process

The Hawkes process (Hawkes 1971) is a self-exciting mathematical model that describes the occurrence of events over time. Its conditional intensity function is defined as:

$$\lambda(t) = \mu + \sum_{j:t_j<t} \psi(t - t_j) \tag{1}$$

where μ denotes the basic intensity. $\psi(t - t_j)$ is a prespecified decay function that describes how past events affect future events. Further, neural Hawkes process (Mei and Eisner 2017) generalizes the classical Hawkes process by parameterizing its intensity function with a RNN:

$$\lambda(t) = f(\boldsymbol{w}^\top \boldsymbol{h}(t)),\ f(x) = \beta \log(1 + \exp(\frac{x}{\beta})) \tag{2}$$

where $\boldsymbol{h}(t)$ is the hidden state of the historical event encoded by the continuous-time LSTM, $\boldsymbol{w}$ is the learnable weight, and $f(\cdot)$ is the softplus function with parameter β.

Conditional Intensity Function

Let $\mathcal{H}_k^i = \{(u_j, t_j) | t_j < t_o\} = C_{k,i}(t_o)$ denote the i-th cascade sequence of the topic m_k before the observation time t_o. Inspired by the Hawkes process, we innovatively define the conditional intensity function of the i-th cascade sequence of topic m_k at time t as:

$$\lambda_k^i(t|\boldsymbol{\theta}, \mathcal{H}_k^i) = f(\underbrace{\sum_{j:t_j<t} g(t - t_j) \cdot \boldsymbol{w}_h^\top \boldsymbol{h}(t_j)}_{\text{history}} + \underbrace{\boldsymbol{w}_\Phi^\top \boldsymbol{\Phi}(t)}_{\text{intra-topic}} + \underbrace{\boldsymbol{w}_\Psi^\top \boldsymbol{\Psi}(t)}_{\text{inter-topic}} + \underbrace{\boldsymbol{w}_b^\top \boldsymbol{x}_{m_k}}_{\text{base}}) \tag{3}$$

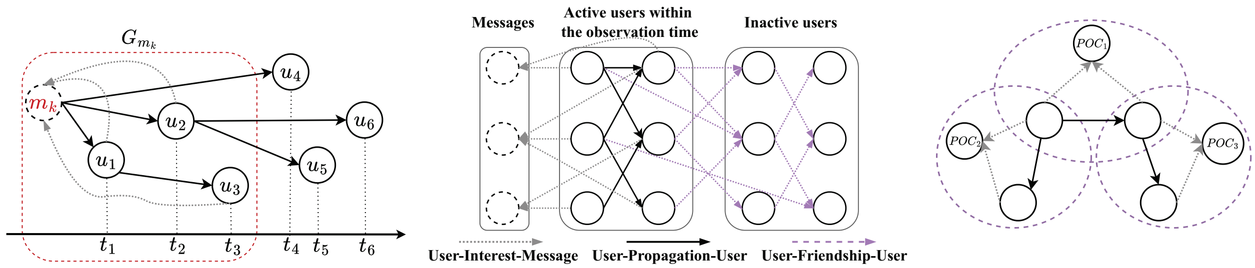
where $\boldsymbol{\theta}$ is the set of trainable parameters. Eq. (3) shows that the conditional intensity function consists of the following four parts:

(1) **history.** The "*history*" term is used to measure the influence of retweeting behavior occurring earlier in the sequence on the conditional intensity at future moments. We use GRU (Cho et al. 2014) to encode the user who retweeted at time t_j as $\boldsymbol{h}(t_j)$. $g(t - t_j) = e^{-\beta(t-t_j)}$ denotes the influence of the retweeting behavior occurring at t_j on the conditional intensity at t.

(2) **intra-topic.** The "*intra-topic*" term measures the influence of other cascade sequences of the same topic on the current cascade sequence. For example, in Fig. 1(a), u_6 may be influenced by his/her friend u_4 to retweet m_k. We use the hidden state at the last moment in other cascade sequences to represent the influence of "*intra-topic*", and it decreases over time:

$$\boldsymbol{\Phi}(t) = \sum_{c=1 \wedge c \neq i}^{R_k} g(t - t_c) \cdot \boldsymbol{h}(t_c) \tag{4}$$

where t_c is the last retweeting time in the c-th cascade sequence $\mathcal{H}_k^c$.



(a) An example of information cascade. (b) Heterogeneous cascade graph. (c) An example of several POFs.

Figure 1: Related illustrations of public opinion field effect.

(3) **inter-topic.** The "*inter-topic*" term is used to measure the mutual influence of different topics. For example, the basketball enthusiast u_2 in Fig. 1(a) has retweeted another basketball topic m_b before t_2, and the retweeting of m_b drives u_2 to retweet m_k. We use $\boldsymbol{\Psi}(t)$ to measure this influence:

$$\boldsymbol{\Psi}(t) = \sum_{b=1 \wedge b \neq k}^{|M|} g(t - t_b) \cdot \boldsymbol{x}_{m_b} \tag{5}$$

where t_b denotes the time when topic m_b was first retweeted and $\boldsymbol{x}_{m_b}$ denotes the embedding of topic m_b. We also use a time decay function to attenuate the influence of topics that are propagated for a longer period of time.

(4) **base.** The "*base*" term denotes the basic intensity, represented by the initial feature $\boldsymbol{w}_b^\top \boldsymbol{x}_{m_k}$ of the topic node m_k.

Finally, to ensure that the intensity is positive while avoiding drastic changes in intensity, we use the softplus function $f(\cdot)$ to map the sum of the above four terms.

Maximum Likelihood Estimation

The conditional intensity function of the i-th cascade sequence $\mathcal{H}_k^i$ of the topic m_k at time t is $\lambda_k^i(t|\boldsymbol{\theta}, \mathcal{H}_k^i)$, and its time window is $[t_1, t_2, ..., t_o)$. According to $\lambda_k^i(t|\boldsymbol{\theta}, \mathcal{H}_k^i)$, the probability of observing a retweeting event in the cascade sequence at time t ($t > t_o$) after the observation time t_o is defined as:

$$P(t) = \lambda_k^i(t) \exp(-\int_{t_o}^{t} \lambda_k^i(\tau) d\tau) \tag{6}$$

where $\tau \in [t_o, t]$. Our goal is to maximize the log-likelihood $\ell_{k,i}$ of observing each user's retweeting event in $\mathcal{H}_k^i$:

$$\begin{aligned}\ell_{k,i}(\boldsymbol{\theta}|\mathcal{H}_k^i) &= \sum_{j:t_j<t_o} \log(P(t_j)) \\ &= \underbrace{\sum_{j:t_j<t_o} \log(\lambda_k^i(t_j|\boldsymbol{\theta}))}_{\text{retweet}} - \underbrace{\int_{t=t_1}^{t_o} \lambda(t) dt}_{\text{non-retweet}}\end{aligned} \tag{7}$$

where $\lambda(t) = \sum_{k=1}^{|M|} \sum_{i=1}^{R_k} \lambda_k^i$ denotes the sum of the intensity of all cascades in the entire heterogeneous cascade graph. The first term of Eq. (7) is the probability of retweeting at each time in the cascade sequence; the second term represents the probability of not retweeting in an infinitely small interval $[t, t + \Delta t)$. Since the probability of not existing any retweeting event within the interval $[t, t + \Delta t)$ is $1 - \lambda(t)dt$, its log is $-\lambda(t)dt$.

Ultimately, the loss function is defined as the opposite of the sum of the log-likelihood functions of all cascade sequences:

$$\mathcal{L}_m = -\sum_{k=1}^{|M|} \sum_{i=1}^{R_k} \ell_{k,i} \tag{8}$$

Public Opinion Field

(Li et al. 2023) proposes that every public opinion field has a center that dominates it, namely "*public opinion center node*" (POC node). It is obvious that in a HCG, each topic node can be considered as the "*whirlpool center*" of the corresponding cascade. With a topic at the center, the cascade spreads among users. Next, we define the public opinion field as follows.

Definition 4 (Public Opinion Field, POF)**.** *Given a heterogeneous cascade graph $G_{t_o} = (\mathcal{V}, \mathcal{E}_{t_o}) = (U \cup M, E \cup D_t \cup I_t)$ and a topic $m_k \in M$, the public opinion field P_k centered on POC node m_k is defined as a subgraph $G_{m_k} = (\mathcal{V}_{m_k}, \mathcal{E}_{m_k})$ of G_{t_o}. $\mathcal{V}_{m_k} = \{m_k\} \cup U_{m_k}$ denotes the set of nodes, and U_{m_k} denotes the user who has retweeted m_k before observation time t_o. $\mathcal{E}_{m_k} \subseteq (D_t \cup I_t)$ denotes the set of edges, i.e., the propagation relations between U_{m_k} and the interest relations between U_{m_k} and m_k.*

A public opinion field G_{m_k} formed with m_k at the center at observation time t_3 is given in Fig. 1(a). Fig. 1(c) illustrates an example of several POC nodes and their corresponding POFs in a HCG. In fact, a public opinion field records the propagation of the corresponding topic before the observation time. As described in (Li et al. 2023), there are two well-known real-world public opinion field effects that influence the subsequent propagation of topic:

Observation 1 (***Attention Competition Effect***)**.** *News popularity declines as the number of competing items increases, with a user's attention span remaining constant and limited in the diversity of memes they can focus on (Lorenz-Spreen et al. 2019).*

Observation 2 (***Popularity Dominance Effect***)**.** *In public opinion space, the topics with the higher popularity are more* *powerful to attract users' attention (Schulz and Roessler 2012).*

The two public opinion field effects show that topics compete for user attention during the propagation process, and topics with stronger "*ability*" will attract more attention and be retweeted more. We use "*public opinion field energy*" (POFE) proposed in (Li et al. 2023) to quantify the ability of a topic to attract attention. The greater the conditional intensity, the greater the probability of a future retweeting event. Retweeting behavior is the result of the success of a topic in grabbing the user's attention, so conditional intensity and public opinion field energy are naturally linked. In this paper, we use the sum of the conditional intensity of the last time in all cascade sequences in a POF as the POFE of this POF. Specifically, the energy of the public opinion field P_k for topic m_k is defined as:

$$E(P_k) = \sum_{i=1}^{R_k} \lambda_k^i(t_i) \tag{9}$$

where t_i is the last retweeting time in the i-th cascade sequence $\mathcal{H}_k^i$.

Representation Learning

This section explores the combination of public opinion field effect with HCG's node representation learning for information popularity prediction. Specifically, the representation learning is divided into three modules: (1) Public opinion field learning. Learning the representation of POC nodes, i.e., topic nodes; (2) User global dependency learning. Learning user representation through user propagation, interest and friendship relationships; (3) Prediction module. Pooling the public opinion field based on the node's representation learned earlier, and then performing information popularity prediction. We will describe each module in detail below.

Public Opinion Field Learning

Because the nodes with different types always have different feature spaces, we first utilize $\boldsymbol{h}_i = \boldsymbol{A}\boldsymbol{x}_i$ to project the features of all the nodes in $G_{t_o} = (\mathcal{V}, \mathcal{E}_{t_o})$ into the same space, where $\boldsymbol{x}_i$ is original feature of node $v_i \in \mathcal{V}$, $\boldsymbol{A}$ is a projection matrix, and $\boldsymbol{h}_i \in \mathbb{R}^d$ is the projected features. Note that $\boldsymbol{A}$ is different depending on the type of node, i.e., U and M.

Consider that the energy of different POFs has different scales, to simplify the comparison, we first normalize the energies of all POFs to between 0 and 1:

$$E(P_k) \leftarrow \frac{E(P_k)}{\sum_{m_k \in M} E(P_k)} \tag{10}$$

In this paper, we use the popular graph attention mechanism (Velickovic et al. 2017) to learn the dependencies between a topic and its user neighbors. Specifically, the attention score of node $v_i \in \mathcal{V}$ and its neighbor $v_j \in \mathcal{V}$ is:

$$\mathrm{att}(\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{\Phi}) = \sigma(\boldsymbol{\Phi}^\top \cdot (\boldsymbol{h}_i \| \boldsymbol{h}_j)) \tag{11}$$

where σ denotes the LeakyReLU activation function, $\boldsymbol{\Phi} \in \mathbb{R}^{2d}$ denotes the weight to perform the attention mechanism, and $\|$ denotes the concatenation operation.

Next, given a topic m_k, its node representation $\boldsymbol{h}_{k,1}$ based on "*user-interest-topic*" relationship is shown as follows:

$$\boldsymbol{h}_{k,1} = \sum_{v_i \in \mathcal{N}_{k,1}} \alpha_{k,i} \cdot \boldsymbol{h}_i \tag{12}$$

where $\mathcal{N}_{k,1}$ denotes the set of user neighbors of m_k. $\alpha_{k,i}$ denotes the attention score between m_k and its neighbor $v_i \in U$, with the following expression:

$$\alpha_{k,i} = \frac{\exp(g(t_o - t_i) \cdot \mathrm{att}(\boldsymbol{h}_k, \boldsymbol{h}_i, \boldsymbol{\Phi}_o))}{\sum_{v_j \in \mathcal{N}_{k,1}} \exp(g(t_o - t_j) \cdot \mathrm{att}(\boldsymbol{h}_k, \boldsymbol{h}_j, \boldsymbol{\Phi}_o))} \tag{13}$$

where $g(t_o - t_i)$ is used to describe the different importance of retweeting behaviors at different times. $\boldsymbol{\Phi}_o$ is a learnable parameter and t_i denotes the time when user v_i retweets m_k.

User Global Dependency Learning

In information popularity prediction, users are the pivotal participants in the information propagation, so it is crucial to consider the dependency relationships among users and between users and topics. In a HCG, user dependencies include "*user-propagation-user*", "*user-interest-topic*" and "*user-friendship-user*" relationships.

User-Propagation-User. There is a natural temporality in the propagation relationship between users. This temporality is usually categorized into short-term and long-term dependence.

(1) Short-term dependence usually refers to the phenomenon of immediate reaction in information propagation. For example, users engage in retweeting behaviors shortly after receiving new topics, and such behaviors are often directly influenced by the most recently exposed users. We use the graph attention mechanism to extract associations between users and the nearest users around them to reflect the short-term dependence of propagation. Specifically, the node representation $\boldsymbol{h}_{i,1}$ of user $v_i \in U$ based on short-term dependencies is:

$$\begin{gathered} \boldsymbol{h}_{i,1} = \sum_{v_j \in \mathcal{N}_{i,1}} \alpha_{i,j} \cdot \boldsymbol{h}_j \\ \alpha_{i,j} = \mathrm{softmax}(\mathrm{att}(\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{\Phi}_s)) \end{gathered} \tag{14}$$

where $\mathcal{N}_{i,1}$ denotes the neighbors (including itself) of user v_i based on the propagation relationship, $\alpha_{i,j}$ denotes the attention score, and $\boldsymbol{\Phi}_s$ is the learnable weight.

(2) Long-term dependence involves the cumulative effect of information propagation, which focuses on how behavior at a more distant time point in the past affects the current user's decisions. Transformer (Vaswani et al. 2017) is able to capture complex dependencies in long-distance sequences. We first form a user propagation matrix $\Omega \in \mathbb{R}^{R \times N}$ of all cascade sequences of all topics:

$$\Omega = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ m_k & u_i & u_j & \cdots \\ m_{k+1} & u_s & u_t & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}_{R \times N} \tag{15}$$

where $R = \sum_{k=1}^{|M|} \sum_{i=1}^{R_k}$ denotes the number of cascade sequences and N denotes the maximum length of all cascade sequences. To facilitate parallel operations, we use 0 for padding. Therefore, there is a corresponding mask matrix $\Omega_{mask} \in \mathbb{R}^{R\times N}$. $\Omega_{mask}(i,j) = 1$ indicates that a user exists at that location, otherwise $\Omega_{mask}(i,j) = -\inf$ indicates the absence of a user. In order to consider the time at which different users are involved in retweeting, we define the time encoding matrix $\Omega_{time} \in \mathbb{R}^{R\times N}$. $\Omega_{time}(i,j)$ denotes the value after normalizing the Unix timestamps of the user's retweeted topic at position (i,j). Then, the Transformer operation is defined as follows:

$$\begin{aligned} \boldsymbol{Q} &= \boldsymbol{W}_Q(\boldsymbol{H} + f_t(\Omega_{time})) \\ \boldsymbol{K} &= \boldsymbol{W}_K(\boldsymbol{H} + f_t(\Omega_{time})) \\ \boldsymbol{V} &= \boldsymbol{W}_V(\boldsymbol{H} + f_t(\Omega_{time})) \\ \boldsymbol{H}^{'} &= \mathrm{softmax}((\boldsymbol{Q}\boldsymbol{K}^{\top} \odot \tilde{\Omega}_{mask})/\sqrt{d})\boldsymbol{V} \end{aligned} \tag{16}$$

where $\boldsymbol{H} \in \mathbb{R}^{R\times N\times d}$ denotes the embedding matrix of all entities in Ω. $\boldsymbol{W}_Q$, $\boldsymbol{W}_K$ and $\boldsymbol{W}_V$ are the learnable parameters. $f_t(\cdot)$ denotes the time encoding function, a simple MLP which maps the time encoding matrix Ω_{time} from $\mathbb{R}^{R\times N}$ to $\mathbb{R}^{R\times N\times d}$. We broadcast $\Omega_{mask} \in \mathbb{R}^{R\times N}$ to $\tilde{\Omega}_{mask} \in \mathbb{R}^{R\times N\times N}$ to fit the shape of the attention matrix. Finally, we obtain the representation matrix $\boldsymbol{H}^{'} \in \mathbb{R}^{R\times N\times d}$ for all entities after extracting the long-term dependencies. Since our task is to predict topic nodes, we regard the first column in $\boldsymbol{H}^{'}$ as the calculation result of long-term dependency. Considering that any $m_k \in M$ may have multiple cascade sequences, we average the R_k sequence representations of m_k to obtain its final representation, that is, the representation $\boldsymbol{h}_{k,2}$ of topic m_k based on long-term dependencies:

$$\boldsymbol{h}_{k,2} = \frac{1}{R_k} \sum_{\Omega[i,0]=m_k} \boldsymbol{H}^{'}[i,0,:] \tag{17}$$

User-Interest-Topic. Users' interest in the topic directly determines whether they will participate in further propagation of the topic, and thus providing useful knowledge for subsequent popularity prediction. In general, a user usually retweets multiple topics. For each user node, its different POC neighbor nodes (topics) will compete for user v_i's attention (***Attention Competition Effect***). The POC neigbor node with the larger POFE will gain more attention from v_i (***Popularity Dominance Effect***). We consider both self attention mechanism and public opinion field effect in node representation $\boldsymbol{h}_{i,2}$ for user v_i:

$$\begin{aligned} \boldsymbol{h}_{i,2} &= \sum_{m_k\in\mathcal{N}_{i,2}} \alpha_{i,k}\cdot(\boldsymbol{h}_k + E(P_k)\odot \boldsymbol{h}_k^{'}) \\ \alpha_{i,k} &= \mathrm{softmax}(\mathrm{att}(\boldsymbol{h}_i, \boldsymbol{h}_k, \boldsymbol{\Phi}_t)) \end{aligned} \tag{18}$$

where $\mathcal{N}_{i,2}$ denotes the POC neighbors of user v_i based on interest relationship. Note that $\boldsymbol{h}_k^{'} = \boldsymbol{h}_{k,1} + \boldsymbol{h}_{k,2}$ is the final representation of the POC node m_k and $E(P_k) \leftarrow E(P_k)/\sum_{m_j\in\mathcal{N}_{i,2}} E(P_j)$ denotes the normalized public opinion field energy. In Eq. (18), the term $E(P_k)\odot \boldsymbol{h}_k^{'}$ is utilized to reflect the influence of POF P_k for representation of v_i, because $\boldsymbol{h}_k^{'}$ is the representation of P_k.

User-Friendship-User. Since the first two relationships can only last until observation time t_o, neither of them can establish interactions between users in the POF and potential future retweeters. By considering the friend relationship between users, users inside and outside the POF can be associated to build a more comprehensive user portrait. The node representation $\boldsymbol{h}_{i,3}$ of user $v_i \in U$ based on friendship relationship is shown as follows:

$$\begin{aligned} \boldsymbol{h}_{i,3} &= \sum_{v_j\in\mathcal{N}_{i,3}} \alpha_{i,j}\cdot \boldsymbol{h}_j \\ \alpha_{i,j} &= \mathrm{softmax}(\mathrm{att}(\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{\Phi}_f)) \end{aligned} \tag{19}$$

where $\mathcal{N}_{i,3}$ denotes the friends of user v_i (including v_i), $\alpha_{i,j}$ denotes the attention score, and $\boldsymbol{\Phi}_f$ is the weight for performing the attention mechanism.

Ultimately, we get the final representation of user v_i based on the representation of three relations:

$$\boldsymbol{h}_i^{'} = \boldsymbol{h}_{i,1} + \boldsymbol{h}_{i,2} + \boldsymbol{h}_{i,3} \tag{20}$$

Prediction Module

The incremental popularity of a topic depends not only on its intrinsic attributes (e.g., topic heat), but is also influenced by the number of retweeted users (e.g., the number of high-influence users). In other words, the popularity increment is related to the **public opinion field**, as the public opinion field accurately records the spread of the topic up to the observation time.

In this paper, we first use a readout function to aggregate the features of all the nodes within the public opinion field P_k formed by the topic m_k in order to generate the overall embedding $\boldsymbol{e}_k \in \mathbb{R}^d$ of P_k:

$$\boldsymbol{e}_k = f_r(G_{m_k}) \tag{21}$$

where $f_r(\cdot)$ denotes a graph pooling function with the following expression:

$$f_r(G_{m_k}) = \boldsymbol{h}_k^{'} \odot (\sum_{v_i\in U_{m_k}} e_{k,i}\cdot \boldsymbol{h}_i^{'}) \tag{22}$$

where U_{m_k} denotes the user who has retweeted m_k before observation time t_o, $\odot$ denotes the Hadamard product. $\boldsymbol{h}_k^{'}$ and $\boldsymbol{h}_i^{'}$ denote the node representations of topic m_k and user $v_i \in U_{m_k}$, respectively. Since different users within P_k contribute differently to the incremental popularity of topic m_k in the future, we use $e_{k,i}$ to measure the contribution of user $v_i \in U_{m_k}$ to topic m_k with the following expression:

$$\begin{aligned} w_{k,i} = \sigma(\mathrm{att}(\boldsymbol{h}_k^{'}, \boldsymbol{h}_i^{'}, \boldsymbol{\Phi}_p)),\ \tau_{k,i} = \sigma(\mathrm{cov}(\boldsymbol{h}_k^{'}, \boldsymbol{h}_i^{'})) \\ e_{k,i} = \frac{\exp(w_{k,i}\cdot\tau_{k,i})}{\sum_{v_j\in U_{m_k}} \exp(w_{k,j}\cdot\tau_{k,j})} \end{aligned} \tag{23}$$

where $w_{k,i}$ reflects how much popularity or attention can be brought if user v_i participates in the discussion or propagation of topic m_k. It is obvious that Michael Jordan brings more attention than the ordinary person when discussing basketball topics.

$\tau_{k,i}$ reflects how much attention that user v_i intends to pay on the topic m_k. Considering that most people prefer to talk about information that is more relevant to them, we select covariance function $\operatorname{cov}(\boldsymbol{x}, \boldsymbol{y}) = (\sum_{i=1}^{d}(x_i - \bar{\boldsymbol{x}})(y_i - \bar{\boldsymbol{y}}))/(d-1)$, which indicates the correlation between two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, to measure the amount of attention that user v_i intends to pay on the topic m_k.

Then we feed the overall representation $\boldsymbol{e}_k$ of the public opinion field P_k into a MLP to obtain incremental popularity prediction of the cascade C_k:

$$\widehat{\Delta R_{C_k}} = \mathrm{MLP}(\boldsymbol{e}_k) \tag{24}$$

We use the Mean Squared Logarithmic Error (MSLE) as the loss function, which can be formulated as follows:

$$\mathcal{L}_s = \frac{1}{Y}\sum(\log(\Delta R_{C_k}) - \log(\widehat{\Delta R_{C_k}}))^2 \tag{25}$$

where Y represents the number of training samples. Finally, we define the total loss by linearly combining the supervised loss $\mathcal{L}_s$ and the maximum likelihood estimation loss $\mathcal{L}_m$: $\mathcal{L} = \mathcal{L}_s + \rho\mathcal{L}_m + \varphi\mathcal{L}_{reg}$, where ρ is a hyperparameter and φL_{reg} is the regularization term to alleviate overfitting.

Experiments

Experimental Setup

Datasets. Detailed statistical information of the datasets is shown in Table 1. (1) **Twitter** (Hodas and Lerman 2014) contains tweets during October 2010 and their paths through users and friendships between users. (2) **Douban** (Zhong et al. 2012) is collected from a social platform where users post updates about the books they read or the movies they watch. We utilize whether users read the same book or not to build friendship networks. (3) **Android and Christianity** (Sankar et al. 2020) are collected from the community Q&A platform StackExchange, and the cascade corresponds to a chronological series of posts associated with the same hashtag. Finally, we use 80% of the cascades in each dataset as training sets and 20% of the cascades as test sets. Details of data processing are given in the ***Appendix***.

Baselines. For simplicity, our model is denoted as POFHP (**P**ublic **O**pinion **F**ield and **H**awkes **P**rocess). We compare POFHP against four categories of information popularity prediction methods. (1) Feature-based mothods: **XGBoost** (Chen and Guestrin 2016) and **MLP**. (2) Hawkes Process-based methods: **SEISMIC** (Zhao et al. 2015) and **DeepHawkes** (Cao et al. 2017). (3) Graph-based methods: **CasCN** (Chen et al. 2019), **MS-HGAT**(Sun et al. 2022) and **CTCP** (Lu et al. 2023). (4) Public Opinion Field-based method: **POFD** (Li et al. 2023).

Evaluation Metrics. We employed three widely-used metrics: Mean Squared Logarithmic Error (MSLE), Mean Absolute Logarithmic Error (MALE), and Symmetric Mean Absolute Percentage Error (SMAPE). The definitions of all evaluation metrics are given in the ***Appendix***.

Datasets	#Cascades	#Sequences	#Users	#Retweets	#Friendship
Twitter	1044	1044	12627	19408	619262
Douban	2594	2594	24926	19638	758310
Android	313	582	9953	9313	48573
Christianity	145	466	2897	6513	35624

Table 1: Statistics of datasets.

Parameter Settings. The feature dimensions of user and topic are both 64, all models are two-layer, with hidden and output layer dimensions of 32 and 64 respectively, and the number of attention heads (if applicable) is 4. In POFHP, we set $\beta = 2.0$, $\rho = 0.01$, and use a single-layer GRU. For all neural network models, the learning rate is 0.001 and the number of training epochs is 200. We run 10 times with the same partition and report the average results.

Performance Comparison

Table 2 reports the performance of the different methods on the four datasets, and some conclusions can be summarized as follows.

(1) Feature-based methods (XGBoost and MLP) performed the worst due to their inability to extract structural and sequence information. (2) Graph-based models exploit both network structure and sequence information, and thus perform better than Hawkes process-based models that only exploit sequences for self-motivated modeling. (3) POFD is second only to our model POFHP in multiple metrics, proving the effectiveness of considering the public opinion field effect in information popularity prediction. (4) Our model POFHP takes into account both the public opinion field effect and sequence modeling, and thus achieves the best performance on all metrics of the four datasets.

Ablation Study

In this section, we study how the various modules of POFHP affect its performance on the Twitter and Douban datasets. Specifically, we create the following model variants: (1) **w/o HP** removes the Hawkes Process, while the $\mathcal{L}_m$ loss function is removed. (2) **w/o SD** removes short-term dependencies from the user-propagation-user relationship. (3) **w/o LD** removes long-term dependencies from the user-propagation-user relationship. (4) **w/o RD** removes the readout function in the prediction module and directly uses a simple two-layer MLP for prediction.

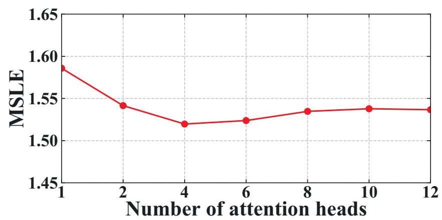
The results of ablation experiments are reported in Table 3. We find that **w/o HP** and **w/o LD** perform poorly, suggesting that it is important to consider complex propagation laws and long-term dependencies. In addition, POFHP outperforms the above four variants, which indicates that our design is reasonable and all modules can improve the performance of the model individually.

Parameter Sensitivity Analysis

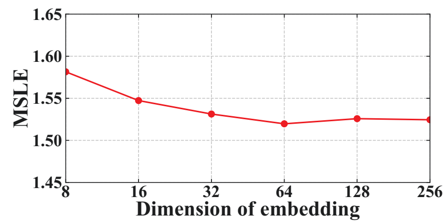
Parameter sensitivity analysis can be used to evaluate how sensitive model performance is to different hyperparameter settings, thus guiding researchers to find the ideal parameter

Methods	Twitter			Douban			Android			Christianity		
	MSLE	MALE	SMAPE	MSLE	MALE	SMAPE	MSLE	MALE	SMAPE	MSLE	MALE	SMAPE
XGBoost	3.2251	1.5726	0.6135	1.8463	1.0853	0.4345	1.7466	0.9951	0.4136	1.4841	0.9127	0.3684
MLP	3.0476	1.5273	0.5965	1.8176	1.1084	0.4454	1.9289	1.0542	0.4416	1.6141	0.9788	0.3895
SEISMIC	2.2377	1.2974	0.5431	1.6160	1.0031	0.4194	1.6151	0.9516	0.3939	1.2614	0.7858	0.3266
DeepHawkes	2.0141	1.2534	0.5285	1.4982	0.9641	0.4025	1.4889	0.9187	0.3789	1.1078	0.6859	0.2987
CasCN	1.9538	1.1623	0.4875	1.5496	0.9788	0.3987	1.3005	0.7674	0.3123	0.8247	0.5515	0.2464
MS-HGAT	2.0539	1.1838	0.4974	1.4718	0.9516	0.4025	1.2274	0.7397	0.2938	0.9015	0.6245	0.2693
CTCP	1.8077	0.9636	0.4578	1.3015	0.8931	0.3776	1.0894	0.6536	0.2775	0.7250	0.5041	0.2257
POFD	1.7774	0.9556	0.4419	1.2387	0.8696	0.3623	1.2834	0.7847	0.3256	0.6045	0.4397	0.2035
POFHP (Ours)	**1.5197**	**0.9149**	**0.4132**	**1.0347**	**0.7342**	**0.3341**	**0.6986**	**0.5208**	**0.2522**	**0.3906**	**0.3676**	**0.1760**
Improvements (%)	↑ 14.50	↑ 4.26	↑ 6.49	↑ 16.47	↑ 15.57	↑ 7.78	↑ 35.87	↑ 20.32	↑ 9.12	↑ 35.38	↑ 16.40	↑ 13.51

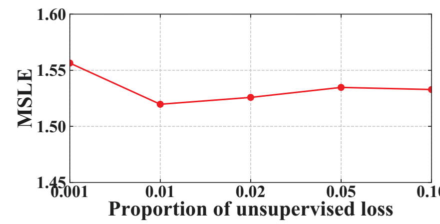
Table 2: Experimental results on 4 datasets. The best results appear in bold.



(a) No. of attention heads.



(b) Hidden dimension of POFHP.



(c) Proportion of unsupervised loss.

Figure 2: Parameter sensitivity analysis of POFHP on Twitter dataset.

Methods	Twitter			Douban		
	MSLE	MALE	SMAPE	MSLE	MALE	SMAPE
w/o HP	1.7012	0.9354	0.4483	1.3258	0.9258	0.3978
w/o SD	1.6113	0.9284	0.4351	1.2478	0.8041	0.3613
w/o LD	1.6514	0.9312	0.4412	1.2784	0.8553	0.3745
w/o RD	1.5438	0.9217	0.4256	1.1517	0.7654	0.3489
POFHP	**1.5197**	**0.9149**	**0.4132**	**1.0347**	**0.7342**	**0.3341**

Table 3: Ablation study of POFHP on Twitter and Douban.

Cases	Target	CTCP	POFD	POFHP
A	3	1.1523	2.8417	2.8963
B	44	42.7936	48.1574	43.8741
C	195	203.7814	193.6587	194.3815

Table 4: Case study experimental results.

combination. We analyze the sensitivity of the parameters in this section and the results are shown in Fig. 2. We observe that the model performance tends to be stable when the number of attention heads, hidden dimension, and unsupervised loss ratio reach 4, 64, and 0.01, respectively. Our model is stable to small parameter changes and performs well under most combinations, demonstrating strong robustness. It ensures reliable results even with slight parameter variations.

Case Study

In order to verify the effectiveness and superiority of POFHP in practical applications, we compare the performance of CTCP, POFD and POFHP on three cases of low popularity (A), medium popularity (B) and high popularity (C) in the Twitter dataset.

We can draw the following three conclusions from the experimental results in Table 4. (1) CTCP is unable to effectively capture the diversion effect of multiple topics on users' attention, and thus its prediction of extreme cases (A and C) is highly biased. (2) POFD considers public opinion field effect but fails to adequately capture the complex laws in information propagation. Thus, POFD is unable to effectively model cases such as B, which has a high initial popularity but not ultimately high popularity. (3) Our POFHP models both public opinion field effects and propagation laws, and can more accurately capture the allocation of users' attention and the complex dynamics of the information propagation process, thus providing more precise prediction results.

Conclusion

In this paper, we propose a novel Hawkes process-based learning model for information popularity prediction. This proposed method first designs an improved neural Hawkes process to capture comprehensive propagation laws from multiple dimensions. In addition, unlike the traditional diffusion model that only considers time dependency and network structure, we propose to use conditional intensity function and cascade diffusion structure to define the public opinion field paradigm and extract high-quality node representations based on the public opinion field effect. Extensive experimental results on four real-world datasets demonstrate the effectiveness of the proposed method.

Acknowledgements

This work was supported by the Key Program of the National Natural Science Foundation of China (Grant No. 62436001), the General Program of the National Natural Science Foundation of China (Grant No. 62472304), and the Joint Funds of the National Natural Science Foundation of China (Grant No. U22A2025).

References

Aravamudan, A.; Zhang, X.; and Anagnostopoulos, G. C. 2023. Anytime user engagement prediction in information cascades for arbitrary observation periods. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 4999–5009.

Cao, Q.; Shen, H.; Cen, K.; Ouyang, W.; and Cheng, X. 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.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chen, X.; Zhou, F.; Zhang, K.; Trajcevski, G.; Zhong, T.; and Zhang, F. 2019. Information diffusion prediction via recurrent cascades convolution. In *2019 IEEE 35th international conference on data engineering (ICDE)*, 770–781. IEEE.

Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Daley, D. J.; Vere-Jones, D.; et al. 2003. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer.

Hawkes, A. G. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1): 83–90.

Hodas, N. O.; and Lerman, K. 2014. The simple rules of social contagion. *Scientific reports*, 4(1): 4343.

Li, C.; Ma, J.; Guo, X.; and Mei, Q. 2017. Deepcas: An end-to-end predictor of information cascades. In *Proceedings of the 26th international conference on World Wide Web*, 577–586.

Li, J.; Yang, Y.; Hu, Q.; Wang, X.; and Gao, H. 2023. Public Opinion Field Effect Fusion in Representation Learning for Trending Topics Diffusion. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Lorenz-Spreen, P.; Mønsted, B. M.; Hövel, P.; and Lehmann, S. 2019. Accelerating dynamics of collective attention. *Nature communications*, 10(1): 1759.

Lu, X.; Ji, S.; Yu, L.; Sun, L.; Du, B.; and Zhu, T. 2023. Continuous-time graph learning for cascade popularity prediction. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 2224–2232.

Masud, S.; Dutta, S.; Makkar, S.; Jain, C.; Goyal, V.; Das, A.; and Chakraborty, T. 2021. Hate is the new infodemic: A topic-aware modeling of hate speech diffusion on twitter. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 504–515. IEEE.

Mei, H.; and Eisner, J. M. 2017. The neural hawkes process: A neurally self-modulating multivariate point process. *Advances in neural information processing systems*, 30.

Sankar, A.; Zhang, X.; Krishnan, A.; and Han, J. 2020. Inf-VAE: A variational autoencoder framework to integrate homophily and influence in diffusion prediction. In *Proceedings of the 13th international conference on web search and data mining*, 510–518.

Schulz, A.; and Roessler, P. 2012. The spiral of silence and the Internet: Selection of online content and the perception of the public opinion climate in computer-mediated communication environments. *International journal of public opinion research*, 24(3): 346–367.

Sun, L.; Rao, Y.; Zhang, X.; Lan, Y.; and Yu, S. 2022. MS-HGAT: memory-enhanced sequential hypergraph attention network for information diffusion prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4156–4164.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *stat*, 1050: 20.

Wang, H.; Dou, Y.; Chen, C.; Sun, L.; Yu, P. S.; and Shu, K. 2023. Attacking fake news detectors via manipulating news social engagement. In *Proceedings of the ACM Web Conference 2023*, 3978–3986.

Wang, J.; Zheng, V. W.; Liu, Z.; and Chang, K. C.-C. 2017. Topological recurrent neural network for diffusion prediction. In *2017 IEEE international conference on data mining (ICDM)*, 475–484. IEEE.

Welling, M.; and Kipf, T. N. 2016. Semi-supervised classification with graph convolutional networks. In *J. International Conference on Learning Representations (ICLR 2017)*.

Yuan, C.; Li, J.; Zhou, W.; Lu, Y.; Zhang, X.; and Hu, S. 2021. DyHGCN: A dynamic heterogeneous graph convolutional network to learn users' dynamic preferences for information diffusion prediction. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part III*, 347–363. Springer.

Zhao, Q.; Erdogdu, M. A.; He, H. Y.; Rajaraman, A.; and Leskovec, J. 2015. 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*, 1513–1522.

Zhong, E.; Fan, W.; Wang, J.; Xiao, L.; and Li, Y. 2012. Comsoc: adaptive transfer of user behaviors over composite social network. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 696–704.