


Research Article

Prediction of New Media Information Dissemination Speed and Scale Effect Based on Large-Scale Graph Neural Network

Chen Qiumeng¹ and Shen Yu ²

¹Minzu Normal University of Xingyi, Guizhou, Xingyi 562400, China

²Huanghe Science and Technology University, Foreign Languages School, Zhengzhou, China

Correspondence should be addressed to Shen Yu; shenyu@hhstu.edu.cn

Received 9 July 2022; Revised 22 July 2022; Accepted 30 August 2022; Published 21 September 2022

Academic Editor: Lianhui Li

Copyright © 2022 Chen Qiumeng and Shen Yu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, because of the popularity of the internet and mobile devices, the dissemination of new media in social networks has attracted extensive attention from scholars and the industry. Scale prediction or propagation speed prediction is to use the initial data to predict the propagation scale of the network. In the complex and changeable social network, how to accurately predict the cascading scale of new media information is the biggest problem at present. In the process of new media information transmission, because of the role of new media information transmission in guiding public opinion, the current hierarchical model of new media information transmission lacks the overall and local models. To solve this problem, a global structure modeling method is proposed. In addition, because of the uncertainty of new media information dissemination, a method of bidirectional recurrent neural network prediction and algorithm complexity is used, and a new method based on large-scale graph neural network is constructed. A prediction method of new media information dissemination speed and scale based on large-scale graph neural network. Through comparative experiments with previous research models, it is found that the NWIDF model constructed in this paper has a good prediction effect.

1. Introduction

In recent years, the modeling and prediction of new media information cascade has attracted extensive attention in the academic field and industry [1]. In recent years, with the improvement of computing power, prediction models based on deep learning have been successful in many tasks.

Existing models based on deep learning can be roughly divided into three categories: (1) models based on information content, such as text, image, video, and other multimedia content, which usually use technology from the field of computer vision and natural language processing to learn the effective representation of information content. (2) Based on time series model, it relies on recurrent neural network, pooling mechanism, and attenuation mechanism to linearly model the information cascade in social networks and [2] (3) model based on the graph structure, such as information cascade graph or global graph. As per reference

[3], these models typically use graph neural networks and graph representation learning techniques to learn efficient structures of nodes, edges, and graphs to represent information. Other deep learning technologies, such as variational reasoning and reinforcement learning, are also used for information cascade scale prediction. In many cases, multimode, multiscale, and multitask learning techniques can be used to improve prediction performance. Deepcas [4] is the first model to model and predict the scale of information cascade using graph representation learning technology. It borrows the idea of deepwalk model [5] and uses the random walk method to sample the information cascade graph. The sampled node sequence is input into the bidirectional gated loop unit [6], and then the node embedding is obtained in cooperation with the attention mechanism [7]. The prediction of the deepcas model is end-to-end, and therefore, it does not depend on the manual functional design. Subsequently, in document [8], the author proposed

the dcgt model, which adds the modeling of node content to deepcas. The purpose of the deephawkes model [9] is to combine the advantages of the generated model with the advantages of deep learning technology, so as to simultaneously consider the predictability and good prediction performance. The ANPP model [10] uses glove [11] for the text embedding of information content and node2vec [12] for user graph embedding. ANPP uses the attention mechanism to aggregate the obtained representations and time-series feature vectors. The Dtcn model [13] predicts the popularity of Flickr images by learning user and image embedding, pore context of shared sequences, and multistep temporal attention mechanism. The dtcn model uses RESNET [14] and long-term and short-term memory artificial neural networks [15] to simulate the visual and temporal dependence of pictures, respectively. The recursive cascade convolution network [16] regards the information cascade graph as a series of subinformation cascade graphs, and then, it uses the dynamic multidirectional graph convolution network to learn the structural information of the information cascade.

Although the model based on deep learning has achieved good results in the information cascade prediction task, it also faces many limitations and challenges. The computational consumption of the deep learning model is generally greater than that of the other two types of models. To obtain satisfactory prediction results, engineers usually need to perform complex parameter optimization and model training and face the risk of data overfitting. At the same time, in the prediction of the cascade scale of new media information, there is a lack of modeling of global and local communication structures, ignoring hierarchical modeling, and it is unable to cope with changes and uncertainties in the process of information dissemination. Therefore, the article starts from this angle, and relevant research is carried out.

2. Related Work

The modeling and prediction of new media information cascades have attracted extensive attention in academia and industry in recent years [1]. In recent years, with the improvement of computing power, prediction models based on deep learning have been successful in many tasks.

Existing deep learning-based models can be roughly divided into three categories: (1) models based on information content, such as text, images, videos, and other multimedia content, these models usually use technologies from the fields of computer vision and natural language processing to effectively represent the content learning of information; (2) time-based models of sequences, which linearly model information cascades in social networks and rely on techniques such as recurrent neural networks, pooling mechanisms, and attention mechanisms [2]. (3) Models based on graph structures, such as information cascade graphs or global graphs, etc. [3], usually use graph neural networks and graph representation learning techniques to learn effective structural representations of nodes, edges, and graphs. Other deep learning techniques, such as variational inference, reinforcement learning, etc., are also

used in information cascade scale prediction. In many cases, techniques, such as multimodal, multiscale, and multitask learning, are used to improve the prediction performance. DeepCas [4] is the first model to use graph representation learning techniques to model and predict the scale of information cascades. It borrows the idea of the DeepWalk model [5] and uses a random walk method to sample the information cascade graph. The sampled node sequence is input into the bidirectional gated recurrent unit [6], and it cooperates with the attention mechanism [7] to obtain the node embedding. The predictions of the DeepCas model are end-to-end and thus do not rely on manual feature design. Subsequently, in literature [8], the authors propose the DCGT model, which adds the modeling of node content to DeepCas. The purpose of the DeepHawkes model [9] is to combine the advantages of generative models with the advantages of deep learning techniques, thereby taking into account both predictive interpretability and good predictive performance. The ANPP model [10] uses GloVe [11] for the textual embedding of information content and node2vec [12] for user graph embedding. ANPP uses an attention mechanism to aggregate the acquired representations and time series feature vectors. The DTCN model [13] predicts the popularity of Flickr images by learning user and image embeddings, sharing temporal context of sequences, and a multistep temporal attention mechanism. The DTCN model uses ResNet [14] and Long Short-Term Memory Artificial Neural Network [15] to model the visual and temporal dependencies of pictures, respectively. Recurrent Cascade Convolutional Networks [16] treat the information cascade graph as a series of subinformation cascade graphs, and then, it use a dynamic multidirectional graph convolutional network to learn the structural information of the information cascade.

Although deep learning-based models have achieved good results on information cascade prediction tasks, they also face many limitations and challenges. The computational consumption of deep learning models is generally larger than that of the other two types of models. The main reason is that deep learning learns the deep nonlinear network structure, and its essence is to approximate complex functions and represent the distributed representation of the input data. Deep learning can learn the essential characteristics of the dataset. But the problem may often involve causal reasoning, logical reasoning, and dealing with uncertainty, which is obviously beyond the ability of traditional deep learning methods. Hence, the predictions of deep learning models lack interpretability, because neural networks are essentially a “black box model.” Secondly, the computational consumption of deep learning models is generally higher than that of feature engineering-based prediction models and probability-based generation. The model should be bigger. To achieve satisfactory prediction results, engineers often need to perform complex parameter tuning, model training, and face the risk of overfitting the data. At the same time, in the prediction of the cascade scale of new media information dissemination, there is a lack of modeling of the global and local dissemination structure, ignoring hierarchical modeling and inability to deal with

changes and uncertainties in the process of information dissemination. Therefore, the paper starts from this perspective, and related research is conducted.

3. Related Theories

3.1. Bayesian Graph Neural Network. A Bayesian network is a probabilistic graphical model. By adjusting the preset parameters or the prior knowledge of the model through sample data, the parameters of the Bayesian network or the posterior probability of the model are inferred to express uncertainty. The uncertainty of the node characteristics of new media information dissemination is mainly manifested in the uncertainty caused by noise data, missing or repeated data, etc., in the process of feature extraction and the uncertainty of the relationship between different characteristics of nodes and node labels. A probabilistic graph model that can solve uncertainty through the Bayesian graph neural network identifies hot topics in new media information, compares the prediction of node labels under different features, integrates the prediction of node labels by all features, and then judges the uncertainty of node features.

In the dissemination of new media information, as the network structure is not fully known and constructed by domain experts, it usually leads to missing important edges, adding false edges and other problems, resulting in poor model prediction effect and poor robustness.

This paper needs to propose a way to add missing important edges and prune irrelevant and spurious edges. In other words, the network structure needs to be reconstructed. The Bayesian graph neural network is used to solve the uncertainty problem of the node relationship in the reconstructed network.

Generally speaking, a neural network can be regarded as a conditional distribution model $p(Y|X, W)$, i.e., the distribution of labels Y under the condition of input feature X and neural network weight W . Then, the learning process of the neural network can be regarded as maximum likelihood estimation. Based on this, the researchers proposed a Bayesian neural network [17], which, firstly, obtained the weight probability $p(W|D)$ of the neural network based on the dataset not only to find its maximum posterior value but also to be used for the neural network. Networks introduce uncertainty. The prediction Y for a new input x can be obtained by integrating the posterior distribution of W , and the process can be expressed as follows:

$$p(y|x, X, Y) = \int p(y|x, W)p(W|X, Y)dW. \quad (1)$$

However, since the posterior distribution (formula (1)) of the Bayesian neural network is often difficult to calculate directly, researchers have adopted different methods to approximate it [18–21].

This paper considers reconstructing the network structure with a random graph generation model to solve the uncertainty of the network structure.

3.2. Random Block Model. The random block model is a generative model for random graphs. The model tends to generate graphs that contain populations, i.e., subsets characterized by a certain edge density interconnected. For example, edges may be more common within a community than between communities. The stochastic block model is important in statistics, machine learning, and network science, and in graph data, it serves as a useful benchmark for the task of recovering community structure. Reconstructing the network structure in this way can aggregate nodes with strong correlations in the network, while nodes with weak correlations will have certain edges if they are directly pressed.

The random block model has the following parameters:

- (1) The number of vertices n
- (2) Divide the vertex set $\{V_1, V_2, \dots, V_n\}$ into disjoint subsets C_1, C_2, \dots, C_r called groups
- (3) A symmetric matrix P of edge probabilities

Randomly sample the edge set: any two nodes $V_p \in C_i$ and $V_q \in C_j$ are connected by an edge with probability P_{ij} .

Its generation process is shown in Algorithm 1.

The group membership of each node depends on the context, i.e., each node may have different memberships when interacting with or being interacted with by different nodes. Statistically, each node is a mixture of group-specific interactions. After the random block model is represented by a generative graph, the network can be reconstructed and applied to the graph neural network to solve the uncertainty of the network structure.

3.3. New Media Information Dissemination Mechanism. In this paper, we mainly explain the mechanism of new media information dissemination from two perspectives, namely information cascade graph and user social network (global graph).

Cascade graph: given the new media information microblog I and its corresponding forwarding information cascade C , the information cascade graph can be defined as $\zeta_c = (v_c, \varepsilon_c)$, where $v_c = \{u_i | 1 < i \leq M\}$ is a part of the user nodes participating in the information cascade, $\varepsilon_c \in v_c \times v_c$ is a set of edges with a number of $M = |C|$, representing all user interactions in the information cascade graph. A schematic diagram of a cascade graph growing over time is shown in Figure 1.

$$C(t_0) \longrightarrow \dots \longrightarrow C(t_n). \quad (2)$$

Global Graph: the global graph contains all the nodes and edges in the social network, which can be defined as $\zeta_g = (v_g, \varepsilon_g)$. The edges represent different node relationships in the information cascade. An example of a typical global graph is the user's follow du and followed network in TikTok.

In this paper, the information cascade graph represents the local propagation characteristics of information in the network, while the global graph represents the associations between nodes in the whole network. Taking TikTok as an

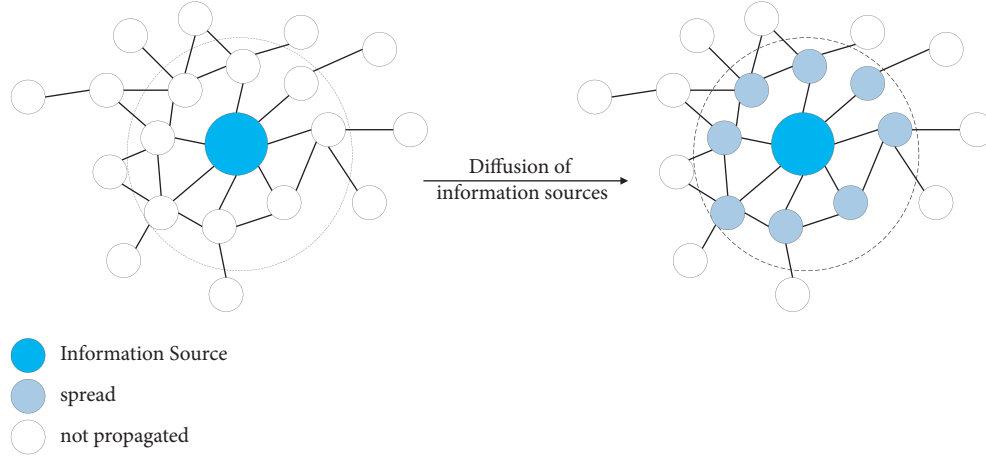


FIGURE 1: Cascading diagram of media information.

example, the following relationships, forwarding relationships, and historical behaviors among users can all be reflected in the structure of the global graph. The previous work [8–10] simply used features, such as the number of followers of the user (which can be regarded as the degree of nodes) as the structural features of the user, which cannot fully capture the user’s influence, preference, and other attributes. There are a few other works [14, 17] that use other types of structural features. But, they also all make strong assumptions about the intrinsic mechanism of information dissemination, or face the risk of overfitting on specific data, resulting in their poor generalization performance, and when migrating to other applications or data platforms (with different propagation mechanism or propagation mechanism unknown) is less effective.

4. New Media Information Dissemination and Scale Prediction Path Based on Large-Scale Graph Neural Network

4.1. The Overall Architecture of the Prediction Model. This section builds the general framework of the NWIDF prediction model. It consists of four parts: structure learning, time series propagation, new media information uncertainty propagation, and predictor. Structural learning mainly captures and models the contextualized structural patterns in information cascade graphs and the implicit relationships of users in social networks. It leverages techniques from graph signal processing to learn structural representations of information cascades: local structure modeling based on wavelet maps and user global structure modeling based on sparse matrix factorization. Temporal propagation uses a bidirectional recurrent neural network to model temporal dependencies in information propagation. The uncertain propagation of information in new media uses a variational autoencoder to model changes and uncertainties in information propagation and information growth, and it uses a regularized flow to estimate the posterior distribution of hidden variables for a series of complex and flexible transformations. The predictor combines recurrent neural

networks and variational inference to learn high-order representations of the information cascade, and finally, it uses a multilayer perceptron to make predictions about the final size of the information cascade, as shown in Figure 2.

As the core of the NWIDF model system, the new media information person is the main body of information production, transmission, processing, and management. Information people usually include are users and platforms. Users can enhance the quality impact of new media platforms by accepting feedback and continuously optimizing. New media information technology is the support of information activities. Through the collection, processing, dissemination, and feedback of information, the continuous operation of the NWIDF model system is realized.

4.2. Modeling of New Media Information Cascade Structure under Large-Scale Graph Neural Network

4.2.1. New Media Information Cascade Learning Structure. In the new media information dissemination mode, the cascade graph is introduced $C_i(t)$, which is represented as an adjacency matrix, and a self-loop is added to each node, as shown in Figure 3. Then, according to the arrival time of each node in the cascade graph $C_i(t)$, one-hot encoding (One-Hot Code) is performed to represent the node characteristics. Divide the observation window $[0, T]$ into disjoint fine l -grained time intervals, then encode each time interval.

At the t_m ($0 < m < l$) moment, the node V_i forwards to the node. Then, V_j the adjacency matrix of the $a_i^{t_m}$ cascaded graph at this moment $C_i(t)$ is 1, and the rest are 0. The adjacency matrix embedding for the cascaded graph $C_i(t)$ is encoded as follows:

$$A_i^T = \{a_1^{t_1}, a_2^{t_2}, \dots, a_i^{t_{i-1}}\}, t_j \in [0, T], j \in [1, l]. \quad (3)$$

To capture the global graph structure information in the process of cascading information diffusion in new media, we use a graph convolutional network to learn the Markov process embedded in [22] information diffusion, i.e., it will converge to a stable value after a period of diffusion. Normal distribution will converge to a stable distribution after a

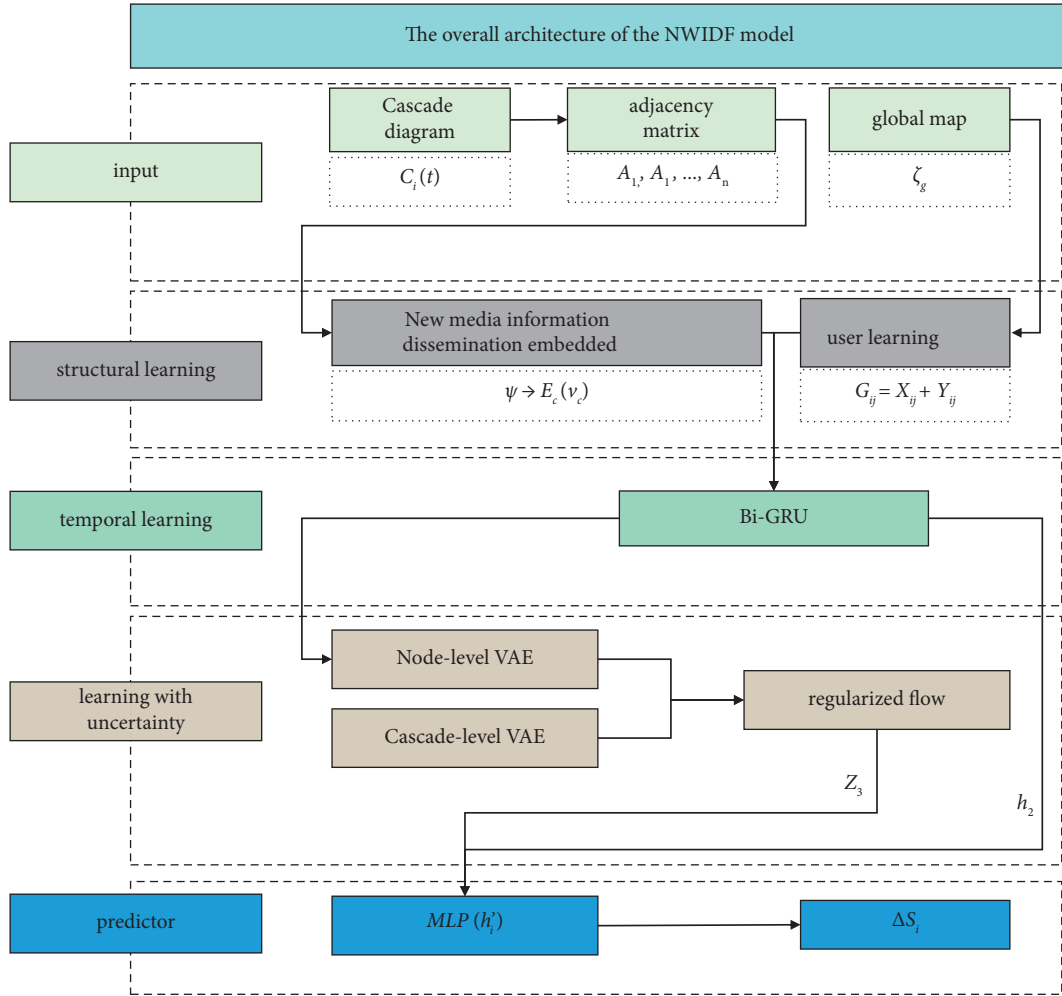


FIGURE 2: The overall architecture of the NWIDF model.

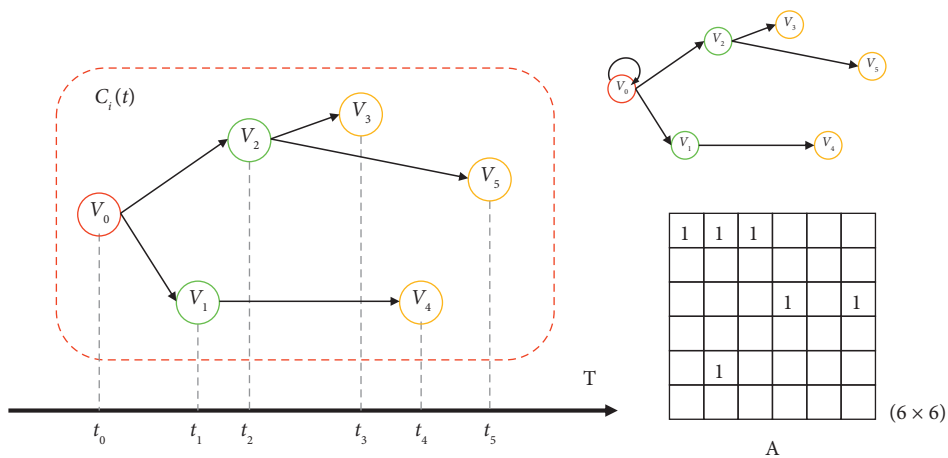


FIGURE 3: Cascaded subgraph sequence sampling representation.

period of time, similar to the normal distribution [23]. Therefore, let the cascaded Laplacian matrix Δc conform to the random walk characteristics of the cascaded graph, and

let the Markov state transition probability matrix $P = D^{-1}A$. According to the graph convolution network formula, the Laplas matrix can be obtained Δc as follows:

$$\Delta c = D^{-1/2}(D - A)D^{-1/2} = D^{-1/2}(I - P)D^{-1/2} \approx \Phi^{1/2}(I - P)\Phi^{-1/2}. \quad (4)$$

D represents the degree matrix of the cascade graph, and K represents the number of captured neighborhood layers.

$\Lambda = \text{Diag}(\lambda_0, \lambda_1, \dots, \lambda_{M-1})$ is the eigenvalue of the adjacency matrix satisfying the condition of $\lambda_0 < \lambda_1 \leq \dots \leq \lambda_{M-1}$, where U is the eigenvalue decomposition. We can then compute $u_i \in v_i(t_0)$ the graph wavelet for each node $\Psi_{u,s}$ as follows:

$$\Psi_{u,s} = U \text{Diag}(g_s(\lambda_0), \dots, g_s(\lambda_{M-1})) U^T \delta_u, \quad (5)$$

where δ_u is the one-hot encoded vector of node u , and the filter kernel g_s is a continuous function defined on R^+ . Here, we use the Heat kernel function $g_s(\lambda) = e^{-\lambda s}$, where s is a scale parameter defined on the spectrum $(\lambda_l)_{l=0, \dots, M-1}$.

In particular, for a given node u_i and a scale parameter s , the empirical feature function is formally defined by the following formula:

$$\varphi_{u,s}(p) = \frac{1}{M} \sum e^{ip\Psi_{m,u,s}}, \Psi_{m,u,s} = \sum_{l=0}^{M-1} g_s(\lambda_l) U_{ml} U_{ul}, \quad (6)$$

where $\Psi_{m,u,s}$ is the m^{th} wavelet coefficient of $\Psi_{u,s}$. Then, the embedding of node u_i in the information cascade graph can be obtained by concatenating the real and imaginary parts.

$$E_c(u_i) = [\text{Re}\varphi_{u,s}(p), \text{Im}\varphi_{u,s}(p)]_{p1, p2, \dots, pd} \quad (7)$$

The dimension of the node embedding $E_c(u_i)$ is $d_c = 2d$, and the first element of the embedding is set as the weight of the node edge, which is defined and regularized by the following formula:

$$W_u = \frac{(t_j - t_0)}{t_0} \in [0, 1], 0 < t_0 \leq t_j. \quad (8)$$

From the perspective of sentences, the emotion of a long sentence is mainly determined by several keywords connected to the root node, and the function of other words is ignored, resulting in the lack of key information. In recent years, the rapid development of Internet+ has made the sentiment analysis of comment texts occupy a certain proportion in user-based big data analysis. Compared with the inflexibility of traditional machine learning methods, deep learning methods can be more efficient and accurate. The emotional information is contained in the text, so obtaining text emotional information through deep learning is a relatively popular research field at present, and it has achieved good research success. Since the machine cannot directly recognize the plain text input, it is necessary to vectorize the text to convert the text into a numerical form that the machine can recognize.

To solve the problem of the adjacency matrix generated by the dependency tree containing a large number of zero elements, there may be information loss and data sparseness. In this paper, a global graph matrix is

constructed, and a layer of identity matrix is added to the original adjacency matrix, which is the global graph matrix. The construction of the global matrix in this paper is shown in equations (11).

$$X_{ii} = 1, \quad (9)$$

$$X_{ij} = 1, \text{ OR } X_{ij} = 0, \quad (10)$$

$$Y_{ij} = 1, \quad (11)$$

$$G_{ij} = X_{ij} + Y_{ij}. \quad (12)$$

$X_{ii} = 1$ means that all elements on the diagonal of the adjacency matrix are set to 1, which means that each node in the graph performs a self-loop operation. The formula $X_{ij} = 1$ indicates that the i^{th} node has a directed connection to the j^{th} node. $X_{ii} = 0$ indicates that there is no connection between the $Y_{ij} i^{\text{th}}$ node and the j^{th} node, and when generating the adjacency matrix, a unit matrix is added, which means that an edge is added between each node in the graph structure to connect. This allows the graph structure to contain global dependency information. G_{ij} represents the connection of all nodes in the graph, which is the global graph matrix. This operation allows each word to play a corresponding role, avoiding data sparse and incomplete information.

Compared with the node embedding in the information cascade graph, the node embedding in the global graph expresses a very different concept of information propagation in the global graph. For the information cascade graph, whether for those influential nodes, hub nodes connecting different communities, or inconspicuous leaf nodes, nodes with similar structural positions will have similar node embeddings even if they are very far apart in the graph. This positional property is captured by the propagation mode of the graph wavelet. For the global graph, the low-dimensional continuous embeddings learned by the model preserve the neighbors of nodes in the global graph. Hence, nodes with similar preferences and behaviors will have similar spatial embeddings.

Unlike information cascade graphs, global graphs often contain up to millions of nodes and edges, making representation learning on them very difficult. Existing graph learning models [15, 18] are difficult to directly apply to practical information cascade prediction problems. The text uses sparse matrix factorization to process and model large-scale global graphs efficiently and in a scalable manner $\zeta_c \zeta_g$.

4.2.2. Temporal Propagation Build. In the above, we used graph wavelets and sparse matrix factorization to generate embeddings that encode the structural information of users in information cascade and global graphs. In particular, they are characterized by the following: (1) structurally equivalent nodes in an information cascade graph will have similar embeddings (refer to [20]). For

example, hub nodes have stronger propagation capabilities than leaf nodes. (2) Adjacent nodes in the global graph will have similar embeddings, i.e., adjacent nodes will have similar preferences for disseminating specific information.

In addition to the structural information contained in the information cascade, time series information is considered to be one of the most important features in the scale prediction problem of the information cascade, and it has a key impact on the final scale of the information cascade. To capture the temporal nature of information cascades, we use bidirectional gated recurrent units (BiGRUs) to model the cascade effects in information. Recurrent neural networks are widely used in the modeling of time series data, and they are used to model time series features in information dissemination. The calculation formula of BiGRU is shown in formulas (12)–(14), which can be expressed in the formula. The hidden state of the forward output at the moment is expressed in watts. The hidden state of the reverse output at time, Q_c , represents the hidden state of the output, ω_c represents the input, and ω_c and v_c represent the weight matrix, where b_c is the bias vector.

$$\vec{Q}_c = GRU(\vec{Q}_{c-1}, \omega_c), \quad (13)$$

$$\bar{Q}_c = GRU(\bar{Q}_{c-1}, \omega_c), \quad (14)$$

$$Q_c = \omega_c \vec{Q}_c + v_c \bar{Q}_{c-1} + b_c. \quad (15)$$

BiGRU includes the process of forward GRU and reverse GRU transfer. Bidirectional GRU can enrich the representation of contextual information based on aspect words and enhance the interaction of information in a complementary form, so that more useful information can be captured compared to unidirectional GRU. Usually, two-way GRU will also perform better than one-way GRU. Deep BiGRU is to continuously expand the depth of the neural network. On the basis of one layer of BiGRU, the method of superimposing multiple layers of BiGRU is to use the output of each BiGRU layer as the input of the corresponding node of the next layer of BiGRU.

However, only using the hidden state of the last layer of the RNN has certain drawbacks for information cascade prediction. It is because of the flat sequence generation process in recurrent neural networks, where the embedding of each node is dependent on the node embedding at the previous time. The problem is that the model is forced to generate all higher-order information in a deterministic and step-by-step manner. This setting has significant limitations for exploring uncertain dependencies in information cascades. In addition, because of the limitations of RNNs themselves, these models cannot handle long-term dependencies, and their predictive performance may drop significantly when the length of the information cascade is very long.

4.2.3. Uncertainty Modeling of New Media Information. An information cascade C consists of a growing sequence of participants, each of which is associated with a learned representation that represents a specific stage of information dissemination. In the above, for each node in the information cascade graph and global graph, we use graph wavelet and sparse matrix factorization to learn its embedding representation $E_c(u_i)$ and $E_g(u_i)$ for the node, respectively. In a more general sense, any other type of graph representation learning method can be used to enhance the learning ability of the model, for example, text and image embeddings. Without causing ambiguity, we use R_i , ($i \in |v_c|$) to represent each participant in the information cascade C , i.e., $R_i = \text{Concat}(E_c(u_i), E_g(u_i))$.

Let $\text{Enc}(\bullet)$ be the input encoder and $\text{Dec}(\bullet)$ be the reconstructed input decoder. The deep variational autoencoder based on neural network can be defined as follows:

$$\begin{aligned} z_i &= \text{Enc}(R_i), \bar{R}_i = \text{Dec}(z_i), \text{ for } i = 1, 2, \dots, M, \\ u_i &= NN(R_i), \log \sigma_i^2 = NN(R_i), z_i \sim N(u_i, \sigma_i^2), \end{aligned} \quad (16)$$

where \bar{R}_i is the reconstructed input and $z_i \in R^d$ is the hidden vector. The variational autoencoder accepts high-dimensional data as input and generates a compressed hidden representation that is sampled from a conditional prior distribution with standard deviation μ and variance $\log \sigma^2$. The original input is then reconstructed from this hidden representation.

In order to learn an efficient probability-based representation from the information cascade data, which captures the variation and uncertainty of the information cascade propagating in the network, the variational autoencoder samples $\log \sigma^2$ and μ and from the output vector of the encoder. Then, use the reparameterization trick to sample the hidden vectors from the Gaussian distribution [22].

$$z_i = u_i + \sigma_i \varepsilon, \varepsilon \sim N(0, 1). \quad (17)$$

Given a hidden random variable $Z \in R^{d_z}$ (in this paper, it is referred to Z_2 learned in higher-order variational autoencoders, the regularization flow is a class of generative models that transforms the observed vector Z into the required target hidden vector Z^k . The transformation consists of a series of K invertible mappings (invertible mappings). The Jacobian matrix of the transformation is computable, and the function is differentiable. In more detail, the regularization flow uses the mapping function $f : Z \rightarrow Z^k$, which is defined as follows:

$$q(Z^k) = q(Z) \left| \det \frac{\partial f^{-1}}{\partial Z^k} \right| = q(Z) \left| \det \frac{\partial f}{\partial Z} \right|^{-1}, \quad (18)$$

where $q(Z)$ is the distribution of the random vector Z and the transfer function f is invertible. To obtain an effective probability density $q_k(Z^{(k)})$ from the initial density $q_0(Z)$, a series of hierarchical transformations of K regularization flows successively use equation (15) to calculate the target density.

Step 1: for each node V_n do
 Step 2: constructing K -dimensional mixed membership vectors $\pi_n \sim \text{Dirichlet}(\alpha)$
 Step 3: for each node pair (V_p, V_q) do
 Step 4: constructor class initialization indicator variable $Z_{p \rightarrow q} \sim \text{Multinomial}(\pi^p)$
 Step 5: construct category indicator variable receiver $Z_{q \rightarrow p} \sim \text{Multinomial}(\pi^q)$
 Step 6: sample their interaction values $Y(p, q) \sim \text{Bernoulli}(Z_{p \rightarrow q} B Z_{q \rightarrow p})$

ALGORITHM 1: Generation process.

$$Z^k = f_k(Z^{(k-1)}) = f_k(f_{k-1}(\dots f_2(f_1(Z))))$$

$$\ln q_k(Z^{(k)}) = \ln q_0(Z) - \sum_{k=1}^k \ln \det \left| \frac{\partial f_k}{\partial Z^k} \right| = q(Z) \left| \det \frac{\partial f}{\partial Z^{(k)}} \right|. \quad (19)$$

If the mapping function is appropriate, then the learned mixture distribution of hidden random vectors more closely matches the distribution of the real data than the simple independent Gaussian distribution.

4.2.4. New Media Information Dissemination Speed and Scale Effect Predictor. Previous studies have found that information cascades have a time decay effect, i.e., the influence of one node on other nodes decreases over time [57]. In this paper, a nonparametric time decay function is used $\lambda_{f(T-t_j)}$, and according to the literature [25], we get the following:

$$h'_i = \sum_{j=1}^{R_i(T)} \lambda_{f(T-t_j)} h'_j. \quad (20)$$

Among them, $f(T-t_j) = \lfloor t_j - t_0/T/l \rfloor$, and R_T indicates the forwarding amount of h'_i new media information, indicating that the time decay function is considered.

The hidden state of the number.

The last part of the NWIDF model is composed of fully connected layers (MLP). According to the previous calculation h'_i , it can be calculated ΔS_i as follows:

$$\Delta S_i = \text{MLP}(h'_i). \quad (21)$$

The final task is to predict the increment of information dissemination within the specified time interval, introducing MeanSquare Log-Transformed Error (MSLE), namely, the following:

$$\text{MSLE} = \frac{1}{N} \sum_{i=1}^N (\log \Delta S_i; -\log \tilde{\Delta S}_i)^2. \quad (22)$$

As the loss function loss, use the Adam optimizer to optimize the loss value to make it optimal (minimum).

Through the above synthesis, the following training process can be performed as shown in Algorithm 2.

5. Experiment Setup and Results Analysis

5.1. Test Setup

5.1.1. Dataset. To evaluate the effectiveness and scalability of NWIDF models in information cascade prediction, experiments are conducted using publicly available datasets and compared with previous studies. The data information statistics of the dataset are shown in Table 1.

Weibo [25]: this dataset selects all the original posts generated by Sina Weibo on June 1, 2016, and tracks all retweets of each post over the next 24 hours, including a total of 119,313 posts. Figure 4(a) shows the distribution of the cascade size; Figure 5(a) shows the prevalence of the cascade, showing that after 24 hours, the prevalence reaches saturation. This paper follows a similar setup to CasCN [26], i.e., observation time window $T=1, 2, 3$ hours. Finally, the stacks are sorted according to the stacking time after preprocessing, and the top 70% of the stacks are selected as the training set for the stacks, and the rest are equally divided into the validation set and the test set.

HEP-PH [27]: the HEP-PH dataset (High Energy Physics Phenomenology Dataset) comes from the electronic version of the arXiv paper citation network. The data covers papers from January 1993 to April 2003 (124 months), in which there are citations for all 34,546 papers. If paper i cites paper j , the paper citation graph contains directed edges from i to j . If a paper cites or is cited by a paper outside the dataset, the graph will not contain information about this. Figure 4(b) shows the distribution of cascade sizes, and Figure 5(b) shows the prevalence of cascades. For the observation window, $T=3, 5,$ and 7 years were chosen, corresponding to the prevalence reaching 50%, 60%, and 70% of the final scale, respectively, as shown in Figure 5(b). Then, 70% of the cascades are collected for training, and the rest are split equally into validation and test sets.

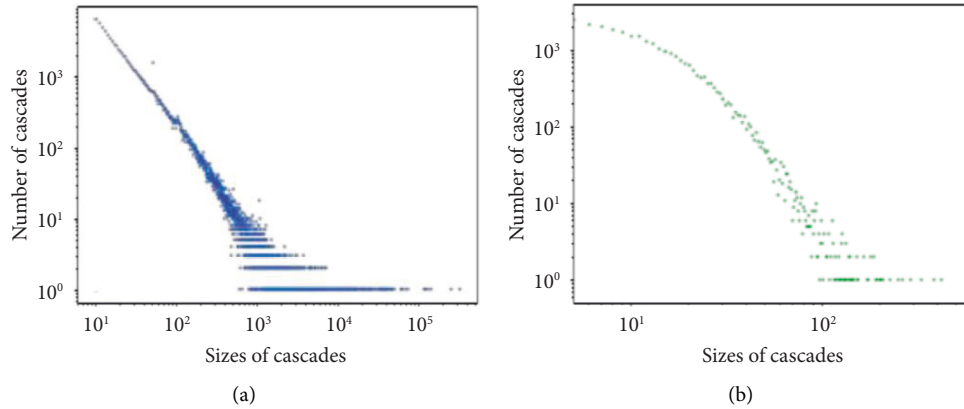


FIGURE 4: Cascade size distribution graph, the X -axis is the cascade size, and the Y -axis is the number corresponding to different cascade sizes. (a) Weibo dataset. (b) HEP-PH dataset.

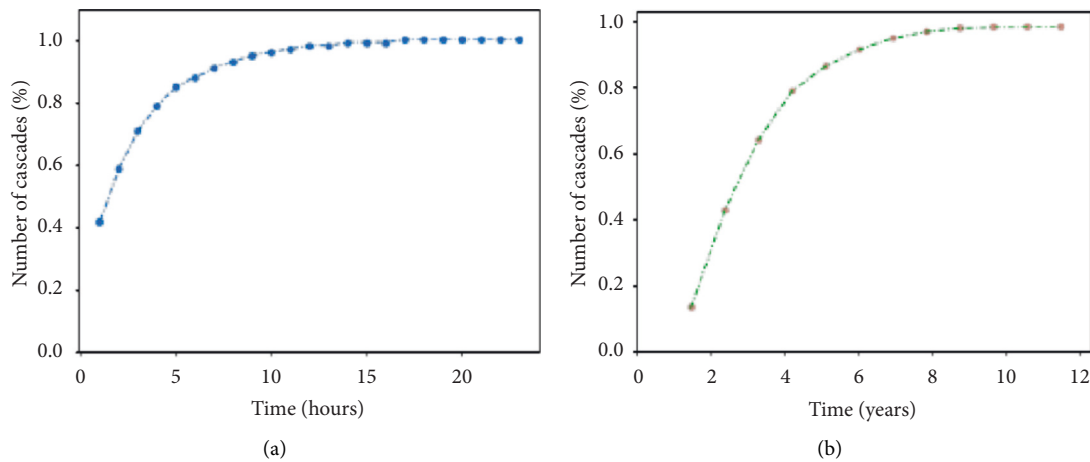


FIGURE 5: Percent distribution between time and number of cascades. (a) Weibo dataset. (b) HEP-PH dataset.

5.1.2. Benchmark Model Selection. From traditional cause analysis methods, hydrological statistics methods, time series analysis methods, etc., to modern artificial neural networks, wavelet theory, gray system and turbidity theory, each method has its own advantages because of its different mechanisms and applicable environments. To verify the effectiveness of our proposed NWIDF model in predicting the scale of information cascades, we choose three basic models: a feature engineering-based model (Topo-LSTM model [29]), a statistical generative model-based DeepHawkes model [25], and a deep learning based model CasCN model [26]. The comparative model is the latest model with high reliability in the research field, which can supplement and improve the comparative analysis research.

5.1.3. Parameter Setting. All experiments in this paper are performed on Ubuntu 16 operating system, Intel Core i9-9980XE CPU, 128G memory, and NVIDIA TiTan RTX (24G) graphics card.

For DeepCas [28], DeepHawkes [25], Topo-LSTM [29], and CasCN [26], refer to DeepCas to set the user's

embedding dimension to 50. The number of hidden units in the fully connected layer of the recurrent neural network is 32 and 16, respectively. The user learning rate 5×10^{-4} , and the other learning rate is 5×10^{-3} . The batch size of each iteration is 32, and when there are 50 consecutive iterations, the loss of the validation set does not drop, and the model training process will stop. The time interval for Weibo dataset was set to 10 minutes, and the time interval for HEP-PH was set to 2 months.

This paper uses Tensorflow to implement the NWIDF model and uses the Adam optimizer to optimize the parameters through gradient descent. Except that, the embedding neighborhood layer of graph representation learning adopts $K=2$, and the rest of the model parameter settings are consistent with the above models.

5.1.4. Evaluation Indicators. According to the existing work, a standard evaluation metric, MSLE (see equation (22)), is selected in the experiment to evaluate the linking accuracy. Note that the smaller the MSLE, the better its prediction performance.

input: cascade graph C , sequence of cascade graph adjacency matrices $A = \{A_1^T, A_2^T, \dots\}$, time window of observation $\Delta c = \{\Delta c_1, \Delta c_2, \dots\}$
output: predicted information cascade incremental scale $\Delta S = \{S_1, S_2, \dots\}$

- (1) the Laplacian matrix of the Δc concatenated graph C ;
- (2) the graph wavelet $\varphi_{u,s}$ for each node u_i ;
- (3) compute the node embeddings of the information cascade graph $E_c(v_c)$;
- (4) Calculate the node embedding of the global graph $E_g(v_g)$;
Calculate the global matrix $G_{ij} = X_{ij} + Y_{ij}$
- (5) while not converge do
- (6) Train a bidirectional gated recurrent unit to acquire h_2 ;
- (7) for $|v_c|$ each user in the pair i do
- (8) calculate z_i
- (9) end for
- (10) get $\vec{Z}_1 = \{z_1, z_1, \dots, z_{|v_c|}\}$;
- (11) Train a cascaded variational autoencoder to obtain Z_2 ;
- (12) Obtained by K transformations Z_3 ;
- (13) Combining sums h_2 and Z_3 sums $\Delta S_i = MLP(h_i)$ to make final scale incremental forecasts;
- (14) end while

ALGORITHM 2: Learning process of the NWIFD model.

5.2. Result Analysis

5.2.1. Experimental Comparative Analysis

(1) *Performance Comparison of the Benchmark Version.* The benchmark version of the NWIFD model proposed in this paper will be experimentally compared with previous cascade prediction models on real datasets.

The DeepCas model [28] is the first deep learning architecture for information cascade prediction, which represents a cascade graph as a set of random walk paths, piped through a bidirectional GRU neural network with an attention mechanism to predict the size of the cascade. It mainly uses the information of structure and node identity for prediction.

The DeepHawkes model [25] integrates the predictive power of end-to-end deep learning into the interpretable factors of the Hawkes process for popularity prediction. The combination between deep learning methods and cascade dynamics modeling processes bridges the gap between the prediction and understanding of information cascades. This method belongs to both generative and deep learning-based methods.

The Topo-LSTM model [29] is a directed acyclic graph structure (DAG structure) RNN that takes a dynamic DAG as input and generates topology-aware embeddings as output for each node in the DAG, thereby predicting the next node.

Pak et al. proposed a particulate matter (PM) prediction model (CNN-LSTM) based on spatiotemporal convolutional network and long short-term memory network and applied it to the concentration prediction of PM2.5 in Beijing. Using mutual information to analyze the spatial-temporal correlation, considering the linear and nonlinear correlation between the target and the observed parameters and combining the historical air quality and meteorological data, the spatial-temporal eigenvectors reflecting the linear

and nonlinear correlation between the parameters are constructed. (STFV), CNN-LSTM prediction model extracts the inherent relationship between PM2.5-related latency air quality and meteorological input data through CNN and reflects the long-term historical process of input time series data through LSTM, using 384 monitoring stations across the country for 3 years. The validity of the model is verified by the air quality data and meteorological data. [26].

Singh et al. used deep learning for stock prediction and proposed a stock prediction model based on 2-dimensional principal component analysis (PCA) and deep neural network (DNN), which combined the closing price, highest price, and lowest price. 36 indicators, such as opening price, are used as the input of the stock prediction model, and the original data matrix is projected into the projection matrix by (2D)2PCA. The dimension of the input sample is reduced, and then the dimension-reduced data is used as the input of DNN in the prediction model. Finally, get the predicted closing price. Compared with the radial basis function neural network (RBFNN), in the stock forecast of Google in Nasdaq, the network (RBFNN) rate is improved by 4.8%, and the actual return (i.e., the correlation coefficient with the predicted return of information dissemination) is 17.1% higher than that of RBFNN [27].

The CasCN model [26] combines the deep learning framework of structure and time, uses graph convolutional network to capture network spatial structure information, and incorporates temporal decay function using a recurrent neural network to achieve the more efficient use of temporal information. This model is a deep learning method.

Table 2 summarizes the performance comparison between the NWIFD model and other model benchmarks on the Weibo and HEP-PH datasets. The comparison of the NWIFD model with the DeepCas model proves that it is not enough to simply embed nodes as a graph representation, and it cannot represent the graph as a set of random paths. Because DeepCas fails to consider timing information and

TABLE 1: Dataset information statistics table.

	Dataset		Weibo_			HEP-PH		
	All		1 hour	2 hours	3 hours	3 years	5 years	7 years
Post-papers edges T	All		119313				34546	
	All		8466858				421578	
		1 hour	2 hours	3 hours	3 years	5 years	7 years	
Cascades	Train	25145	29515	31780	3458	3467	3478	
	Val	5386	6324	6810	837	839	848	
	Test	5386	6324	6810	837	839	848	
Avg.nodes	Train	28.58	29.3	29.48	5.27_	5.27_	5.27_	
	Val	28.71	29.47	29.69	4.32_	4.93_	4.27_	
	Test	29.11	29.77	30.21	4.91_	4.27_	4.28_	
Avg.nodes	Train	27.78	28.54	28.74	4.27_	4.27_	4.27_	
	Val	27.91	28.7	28.94	3.31_	3.93_	3.95_	
	Test	28.32	29.01	29.48	3.91_	3.27_	3.28_	

TABLE 2: Performance comparison table (MSLE).

Data set	Weibo (hours)			HEP-PH (year)		
	1	2	3	3	5	7
DeepCas	2.958	2.689	2.647	1.765	1.538	1.462
Topo-LSTM	2.772	2.643	2.423	1.684	1.653	1.573
Deep-Hawkes_	2.441	2.287	2.252	1.581	1.47	1.233
CasCN	2.242	2.036	1.916	1.004	0.917	0.887
NWIDF	2.123	2.012	1.776	0.939	0.843	0.812

topology of cascaded graphs, its performance is worse than other deep learning-based methods. Topo-LSTM also lacks the processing of timing information, resulting in its poor performance. Although the DeepHawkes models cascades in a generative manner, it does not perform optimally because of its weak ability to learn structural information. CasCN considers temporal information and spatial topology but ignores the fusion between the two features. Finally, the NWIDF model proposed in this paper performs information cascade prediction (tweet retweets and paper citations) on both datasets, which is significantly better than other models. For example, in the Weibo dataset for 1, 2, and 3 hours, the MSLE values were 2.123, 2.012, and 1.776, respectively. Observing in the HEP-PH dataset for 3, 5, and 7 years, the MSLE values are 0.939, 0.843, and 0.812, respectively. The data shows that a good prediction effect has been achieved. Compared with CasCN, the prediction errors of the NWIDF model proposed in this paper are reduced by 5.31%, 1.18%, 7.31%, and 6.47%, 8.07%, 8.46%, respectively, thus confirming the effectiveness of the model.

5.2.2. The Influence of Global Graph on Information Cascade.

NWIDF-All: we removed the structure learning module in the NWIDF model. For NWIDF-All, all nonroot nodes in the information cascade graph are directly connected to the root node, and we do not use global graph information.

Firstly, the validity of the bidirectional recurrent neural network is verified, and experiments are designed. Construct a shortened version of NWIDF-GRU, namely, BiGRU in the benchmark version, and then compare it with CasCN, which is equivalent to NWIDF-GRU, adding attention mechanism on the basis of the CasCN model. During the experiment, the

TABLE 3: Global graph for information cascade prediction performance comparison table (MSLE).

Data set	Weibo (hours)			HEP-PH (year)		
	1	2	3	3	5	7
CasCN $K=2$	2.242_	2.036_	1.916_	1.004_	0.917_	0.887_
NWIDF-All $K=1$	2.136_	2.074	1.858_	1.02_	0.892_	0.842_
NWIDF-All $K=2$	2.099_	2.028_	1.835_	0.92_	0.907_	0.867_

parameters of the two are the same, and the experimental results when the sampling neighborhood layers $K=1, 2$ are shown in Table 3.

The performance comparison of the embedding layer $K=1, 2$ and the CasCN model $K=2$ is given in Table 3. According to Table 3, it can be seen that when $K=2$, the NWIDF model proposed in this paper is better than the CasCN model because the node embedding of the global graph is considered, which can couple the timing information and the spatial structure information. When $K=1$, after observing in Weibo for 2 hours, the MSLE of NWIDF-All is larger than that of CasCN. As $K=1$, the spatial structure information taken is insufficient, resulting in slightly lower results.

Then, the effect of timing information on cascade prediction is verified. Analyze the variant NWIDF-BiGRU of NWIDF, i.e., remove the bidirectional recurrent neural network from the NWIDF model proposed in this paper, and use BiGRU to compare with CasCN, which is equivalent to NWIDF-BiGRU, which is generated by replacing LSTM in the CasCN model with the BiGRU model. From the data in Table 4, it can be seen that when $K=2$, the MSLE = 1.783 observed by Weibo for 3 hours and the MSLE = 0.84 observed by HEP-PH for 7 years are better than the MSLE of CasCN. Thus, the importance of timing information in information cascade is confirmed.

Finally, to verify the impact of time series information and spatial information on cascade prediction, we use BiGRU on the basis of CasCN, add a structure learning mechanism, and then adjust the number of embedded neighborhood layers K to compare with CasCN, respectively. It can be seen from the data in Table 5 that when $K=1$,

TABLE 4: Time series network impact cascade prediction performance comparison table (MSLE).

Data set	Weibo (hours)			HEP-PH (year)		
	1	2	3	3	5	7
CasCN $K=2$	2.242_	2.036_	1.916_	1.004_	0.917_	0.887_
NWIDF-BiGRU $K=1$	2.198_	2.075_	1.872_	0.995_	0.845_	0.884_
NWIDF-BiGRU $K=2$	2.17_	2.065_	1.783_	0.946_	0.826_	0.84_

TABLE 5: Time series space impact cascade prediction performance comparison table (MSLE).

Data set	Weibo (hours)			HEP-PH (year)		
	1	2	3	3	5	7
CasCN $K=2$	2.242_	2.036_	1.916_	1.004_	0.917_	0.887_
NWIDF $K=1$	2.18_	2.032_	1.801_	0.973_	0.819_	0.807_
NWIDF $K=2$	2.123_	2.012_	1.776_	0.939_	0.843_	0.812_

2 on Weibo and HEP-PH datasets, MSLE values are smaller than those of CasCN, indicating that the model proposed in this paper is better than CasCN in terms of time series information and spatial topology information. The capture is more comprehensive, which improves the efficiency of the model and reduces the loss rate.

The above experiments show that time series information and spatial structure information have an important impact on the information cascade prediction effect. The combination of the two can better ensure the accuracy of prediction. The capture of information is also more comprehensive, which also makes the model more generalizable.

6. Conclusion

Under the background of the wrong public opinion orientation caused by the rapid spread of new media information and its wide spread, it is of great practical significance to carry out the prediction of the spread of new media information and the scale effect. Based on the fact that the current hierarchical model of new media information dissemination lacks global and local models, this paper starts from the characteristics that new media information conforms to node dissemination and conducts prediction research on the speed and scale effect of new media information dissemination. The main research contents are as follows [24]:

- (1) The modeling method of local and global propagation of the characteristics of new media information propagation is proposed, and considering the uncertainty and scale effect of information propagation, a graph neural network-based NWIDF model is proposed, which starts from the information cascade structure. Graph the structure learning for information propagation based on locality and globality.

- (2) On the two large-scale information cascade datasets of Facebook and TikTok, the current mainstream and advanced prediction models are applied to carry out experimental research, and it is found that the NWIDF model has better prediction effect and performance. To a certain extent, it can predict the spread and scale of new media information and public opinion, control the rapid spread of wrong public opinion, and quickly cut off the communication channel to provide more advanced ideas.

In this paper, the NWIDF model is the hotspot of current network and graph neural network research. The main difference in speed is the large-scale graph and scale effect of new media information dissemination, and other types of features are caused by the complexity of the research content. In this paper, the NWIDF model is not extended to other types of features, such as learning various content features (number of user attention, h-index of authors, historically published articles, etc.). However, it provides a richer theoretical and practical basis for the future use of more powerful and complex graph neural networks, such as heterogeneous information networks with multiple node types and edge types. The model can be generalized to other types of graph-based business applications, such as virus information diffusion, interpretable information prediction, rumor detection, epidemic control, etc.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

- [1] A. Tatar, M. D. De Amorim, S. Fdida, and P. Antoniadis, "A survey on predicting the popularity of web content," *Journal of Internet Services and Applications*, vol. 5, no. 1, p. 8, 2014.
- [2] Z. Zhang, F. Zhang, and X. Chen, "Etc. Information cascade prediction model based on hierarchical attention," *Computer Science*, vol. 46, no. 6, pp. 210–209, 2020.
- [3] X. Xu, F. Zhou, K. Zhang, S. Liu, and G. Trajcevski, "CasFlow: exploring hierarchical structures and propagation uncertainty for cascade prediction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, p. 1, 2022.
- [4] C. Li, J. Ma, X. Guo, and Q. Mei, "Deepcas: an end-to-end predictor of information cascades," in *Proceedings of the The Web Conference (WWW)*, pp. 577–586, Perth, Australia, April 2017.
- [5] B. Perozzi, R. Al-Rfou, and S. S. Deepwalk, "Online learning of social representations," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, New York, USA, August 2014.
- [6] J. Chung, C. Gulcehre, K. H. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, <https://arxiv.org/abs/1412.3555>.
- [7] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation jointly learning to align and translate," in

- Proceedings of the International Conference on Learning Representations (ICLR)*, pp. 1–15, San Diego, CA, USA, May 2015.
- [8] C. Li, X. Guo, and Q. Mei, “Joint Modeling of Text and Networks for cascade prediction,” *International Conference on Web and Social Media (ICWSM)*, vol. 12, pp. 640–643, 2018.
- [9] G. Chen, Q. Kong, and W. Mao, “An attention-based neural popularity prediction model for social media events,” in *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, pp. 161–163, Beijing, China, July 2017.
- [10] J. Pennington, R. Socher, and C. D. M. Glove, “Global vectors for word representation,” *Conference on Empirical Methods in Natural Language Processing*, vol. 15, pp. 1532–1543, 2014.
- [11] A. Grower and J. L. node2vec, “Scalable Feature Learning for networks,” in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 855–864, San Francisco, California, USA, August 2016.
- [12] B. Wu, W.-H. Cheng, Y. Zhang, and Q. Huang, “Sequential prediction of social media popularity with deep temporal context networks,” in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3062–3068, Melbourne, Australia, August 2017.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, Nevada, USA, June 2016.
- [14] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [15] X. Chen, F. Zhou, K. Zhang, and G. Trajcevski, “Information diffusion prediction via recurrent cascades convolution,” in *Proceedings of the International Conference on Data Engineering (ICDE)*, pp. 770–781, Macao, China, April 2019.
- [16] Q. Cao, H. Shen, J. Gao, and B. Wei, “Popularity prediction on social platforms with coupled graph neural networks,” in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 70–78, Bingzheng Wei, January 2020.
- [17] D. K. Hammond, P. Vandergheynst, and R. Gribonval, “Wavelets on graphs via spectral graph theory,” *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
- [18] C. Donnat, M. Zimik, D. Hallac, and J. Leskovec, “Learning structural node embeddings via diffusion wavelets,” in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 1320–1329, UK, London, July 2018.
- [19] E. Lukacs, *Characteristic functions[M]*, Griffin, London, UK, 1970.
- [20] J. Zhang, Y. Dong, Y. Wang, and J. Tang, “ProNE: fast and scalable network representation learning,” in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4278–4284, Macao, China, August 2019.
- [21] T. Tao, *Topics in Random Matrix theory*, American Mathematical Society, Washington, DC, USA, 2012.
- [22] W. Fan, Y. Ma, Q. Li, and Y. He, *Graph Neural Networks for Social Recommendation*, in *Proceedings of the The World Wide Web Conference*, pp. 417–426, San Francisco, CA, USA, May 2019.
- [23] J. Atwood and D. Towsley, “Diffusion-convolutional neural networks,” *Advances in neural information processing systems*, vol. 29, 2016.
- [24] J. Cheng, L. Adamic, P. A. Dow, and J. M. Kleinberg, “Can cascades be predicted?” in *Proceedings of the The Web Conference (WWW)*, pp. 925–936, Seoul, Korea, April 2014.
- [25] Q. Cao, H. Shen, K. Cen, and O. Wentao, “Deephawkes: bridging the gap between prediction and understanding of information cascades,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1149–1158, Singapore, November 2017.
- [26] X. Chen, F. Zhou, K. Zhang, and T. Goce, “Information diffusion prediction via recurrent cascades convolution,” in *Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 770–781, Macao, China, April 2019.
- [27] J. Leskovec, J. Kleinberg, and C. Faloutsos, “Graphs over time: densification laws, shrinking diameters and possible explanations,” in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 177–187, Chicago, Illinois, USA, August 2005.
- [28] C. Li, J. Ma, X. Guo, and M. Qiao, “Deepcas: an end-to-end predictor of information cascades,” in *Proceedings of the 26th international conference on World Wide Web*, pp. 577–586, Perth, Australia, April 2017.
- [29] J. Wang, V. W. Zheng, Z. Liu, and C. C. C. Kevin, “Topological Recurrent Neural Network for Diffusion Prediction,” in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, pp. 475–484, New Orleans, LA, USA, December 2017.