Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.057714

ARTICLE

Uncovering Causal Relationships for Debiased Repost Prediction Using Deep Generative Models

Wu-Jiu Sun[1] **and Xiao Fan Liu**[1,2,*]

[1]School of Computer Science and Engineering, Southeast University, Nanjing, 211189, China

[2]Department of Media and Communication, City University of Hong Kong, Hong Kong SAR, 999077, China

*Corresponding Author: Xiao Fan Liu. Email: xf.liu@cityu.edu.hk

Received: 26 August 2024 Accepted: 14 November 2024 Published: 19 December 2024

ABSTRACT

Microblogging platforms like X (formerly Twitter) and Sina Weibo have become key channels for spreading information online. Accurately predicting information spread, such as users' reposting activities, is essential for applications including content recommendation and analyzing public sentiment. Current advanced models rely on deep representation learning to extract features from various inputs, such as users' social connections and repost history, to forecast reposting behavior. Nonetheless, these models frequently ignore intrinsic confounding factors, which may cause the models to capture spurious relationships, ultimately impacting prediction performance. To address this limitation, we propose a novel Debiased Reposting Prediction model (DRP). Our model mitigates the influence of confounding variables by incorporating intervention operations from causal inference, enabling it to learn the causal associations between features and user reposting behavior. Specifically, we introduce a memory network within DRP to enhance the model's perception of confounder distributions. This network aggregates and learns confounding information dispersed across different training data batches by optimizing the reconstruction loss. Furthermore, recognizing the challenge of acquiring prior knowledge of causal graphs, which is crucial for causal inference, we develop a causal discovery module within DRP (CD-DRP). This module allows the model to autonomously uncover the causal graph of feature variables by analyzing microblogging data. Experimental results on multiple real-world datasets demonstrate that our proposed method effectively uncovers causal relationships between variables, exhibits strong time efficiency, and outperforms state-of-the-art models in prediction performance (improved by 2.54%) and overfitting reduction (by 7.44%).

KEYWORDS

Repost prediction; causal inference; causal discovery; memory network

1 Introduction

Microblogging platforms enable users to post and share content on their timelines, with 'reposting' functioning as a core mechanism for spreading information online [1]. Accurately predicting repost behavior is therefore valuable across several applications. For instance, advertising and marketing agencies rely on repost prediction to gauge the potential reach of campaigns among target audiences

[2,3]; meanwhile, microblogging services utilize repost likelihood predictions to optimize content recommendations and improve user retention [4]. Recent repost prediction models employ deep learning techniques to extract features from microblogging data, including contents and social relationships, and utilize these features to predict users' repost behaviors based on their associations [5–7].

However, these models often fail to account for confounding variables, which may lead to the learning of spurious associations in the training data, ultimately hindering generalization. For example, as shown in Fig. 1a, news media accounts with large follower bases exert significant social influence, making their content more likely to be reposted by users (social relationship → user behavior) [8]. As a result, once trained, a model may erroneously assume that content from these influential users automatically aligns with the interests of other users. This can lead to inaccurate predictions when the model encounters posts from less influential users (see Fig. 1b). This phenomenon is illustrated in Fig. 1c. Assume S, T, and Y represent social relationships, post content, and user reposting behavior, respectively. Since users receive posts mainly from accounts they follow ($S \rightarrow T$), S acts as a common parent node (confounding variable) for both T and Y. The influence of S can distort the model's understanding of the relationship $T \rightarrow Y$, thereby diminishing the model's prediction performance [9,10].

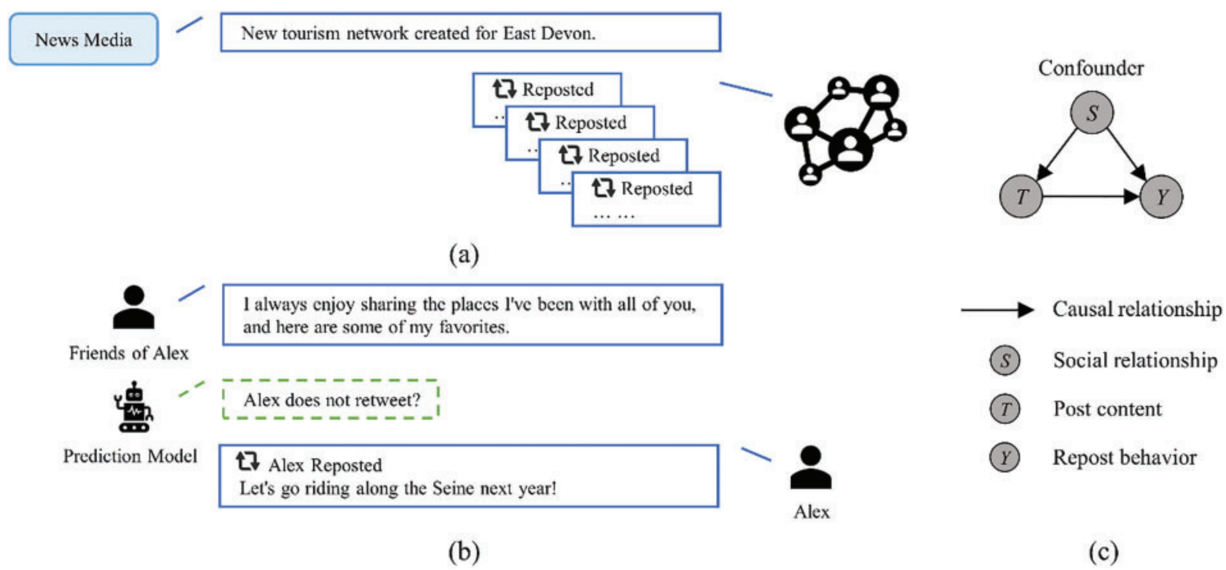

Figure 1: An example of the repost scenario

To address the challenge of confounding variables, recent research has combined deep learning models with causal inference methods. For instance, image recognition models [11–13] utilize causal inference techniques to mitigate the effects of confounders like text and image context. Similarly, recommendation algorithms [14–16] use causal inference to mitigate bias resulting from item popularity. In this context, repost prediction models can leverage causal inference to estimate the causal association between features and user behavior, enabling them to better understand the data and achieve Debiased Repost Prediction (DRP). However, this task poses several challenges from different perspectives. Firstly, identifying confounding variables necessitates the model to possess complete prior knowledge of the variable causality, which may be difficult to achieve in real-world scenarios. Secondly, controlling for the influence of confounding variables requires the model to be aware of the distribution of these variables. In the absence of identifying confounding variables, it is difficult to manually define a confounder dictionary to represent their distribution. We will formally formulate these challenges in Section 3.2.

To address the above challenges, this paper proposes a novel debiased repost prediction model, namely, Causal Discovery for Debiased Repost Prediction (CD-DRP). The proposed method enables the DRP model without assuming a causal relationship. Specifically, the CD-DRP model devises a deep generative network that includes a parameter matrix and several parallel multilayer perceptrons (MLPs). The parameter matrix is the adjacency matrix of the causal graph, which reflects the variable dependencies, while the MLPs are responsible for generating conditional probability distributions between variables. This network aims to identify a causal graph that maximizes the likelihood of microblogging data. This generative network ensures that the proposed model can identify the most explanatory causal graph for the microblogging service. Meanwhile, to facilitate learning the distribution of confounding variables, the CD-DRP model designs a confounder memory network to adeptly capture the information pertaining to these variables within individual data batches. This network can retrieve its memory for various data batches efficiently and works towards minimizing the disparity between the stored memory and the observed confounding variables throughout the training process.

The main contributions of this paper are as follows:

- The CD-DRP model is the first DRP model that enables both causal discovery and debiased repost prediction. It can control the impact of confounders and enhance model generalizability without prior knowledge of the causal graph.
- The CD-DRP model proposes a deep generative network for inferring causal relationships. This network can discover the causal graph that reveals the generation mechanism underlying microblogging data.
- To perceive the distribution of confounding variables, the CD-DRP model designs a confounder memory network. This network can gradually absorb the information of confounders scattered in different data batches by optimizing the reconstruction loss.
- Experimental results on real-world datasets demonstrate that the CD-DRP model effectively captures the causal relationships among variables in the reposting prediction scenario and outperforms state-of-the-art models in predictive performance.

The rest of the paper is arranged as below: Section 2 reviews the related research on causal deep learning and repost prediction. Section 3 provides the necessary background information. Section 4 introduces the CD-DRP model. Section 5 outlines the experimental setup and offers a multi-faceted analysis of experiments, including generalizability, causal graph evaluation, ablation studies, hyperparameter sensitivity, and model efficiency. Finally, Section 6 concludes the paper.

2 Related Work

2.1 Causal Deep Learning

Deep learning is widely used in areas such as image recognition, natural language processing, and recommendation systems with its powerful learning and representation capabilities [17,18]. However, its susceptibility to learn spurious relations affected by confounding variables has been plaguing researchers in the field of deep learning [19]. Since causal inference [20] can control the impact of confounders and evaluate the causal effect of feature variables on predicted targets by do-calculus, i.e., $P(Y|\mathrm{do}(X))$. Recently, there is an increasing number of studies introducing causal inference into deep learning models to implement causal deep learning [11–16]. For example, Yang et al. [11]

improved the attention mechanism based on the front-door criteria of causal inference to help visual language models control the influence of confounding association. Liu et al. [13] developed the CMCIR framework, which leverages causal reasoning to learn cross-modal information, enhancing the framework's ability to understand the causal, logical, and spatiotemporal dynamics between video and language content. This approach effectively addresses event-level visual question answering tasks that demand complex reasoning. Some studies [14,15] have incorporated causal inference to assess the impact of item popularity on recommendations within recommender systems. These frameworks help to reduce popularity bias, ensuring that recommendations rely solely on user and product characteristics.

2.2 Repost Prediction

Predicting reposts plays a key role in opinion analysis and recommendation systems [4,21]. The majority of repost prediction methods concentrate on using machine learning models to identify and learn relevant features. For instance, Jiang et al. [22] extended the probabilistic matrix factorization method by introducing two additional matrices: a social influence matrix based on social network structure and interaction history, and a message similarity matrix based on document semantics. By optimizing the latent feature space of users and messages, the model separately learns social influence and message semantic features, thereby enhancing its predictive performance on user reposting behavior. Safari et al. [23] designed a V-DBNC model based on a dynamic Bayesian network, which can predict users' reposting behavior by learning user behavior patterns, reposting path structure, and social influence between users. In recent years, the development of deep learning technology has highlighted the significant advantages of neural networks over traditional machine learning methods in feature representation. Consequently, neural network-based models for predicting user reposting behavior have gained increasing popularity [5,7]. For example, Wang et al. [24] proposed a dual autoencoder model to capture the features of user identity information, social relationships, and group reposting factors. The model also incorporates an attention mechanism to extract topic representations from users' historical reposting activities, ultimately predicting their behavior by learning these multi-dimensional features. Literature [25] introduced the GODEN model, which combines ordinary differential equations with graph neural networks to capture both dynamic user interactions and static user relationships. The model represents the reposting propagation pattern through user and time context, employing a multi-head attention module to focus on various contextual information and predict the next user likely to forward the content.

2.3 Summary

Existing repost prediction models often neglect confounding variables, which can result in capturing spurious associations between feature variables and user behavior. This limitation hampers the generalization ability of these models. To address this issue and achieve Debiased Repost Prediction (DRP), it is essential to combine repost prediction models with causal inference techniques. However, causal inference typically relies on prior knowledge of the causal relationships, which is challenging to obtain in the context of real-world applications like repost prediction. Correspondingly, due to the lack of prior knowledge of causality, we cannot predefine a confounding variable dictionary [11,13,26] to help the DRP model learn causal associations. Therefore, in this paper, we aim to empower the DRP model to discover variable causality and to learn causal associations without predefining a confounder dictionary.

3 Preliminaries

3.1 Causal Discovery

Recent causal discovery algorithms [27–29] are capable of identifying the structure of a causal Bayesian network from datasets containing n observations and p variables. In this context, a causal Bayesian network is represented as a Directed Acyclic Graph (DAG) with p nodes, where each node corresponds to a variable, and directed edges indicate causal relationships between variables. These algorithms aim to determine the causal Bayesian network that best fits the data, while adhering to the structural constraints of a DAG. We can formulate it as:

$$\min \mathcal{L} = \mathcal{L}_{rec} + \alpha\mathcal{L}_{DAG} + \beta\mathcal{L}_{s}, \tag{1}$$

where the term $\mathcal{L}_{\text{rec}}$ measures the discrepancy between the observed distribution of variables and the distribution of variables generated by the causal Bayesian network. This term reflects the causal Bayesian network's ability to explain observable data. The term $\mathcal{L}_{\text{DAG}}$ quantifies the difference between the causal Bayesian network structure and the DAG structure, ensuring that no causal loop such as $\mathrm{X} \rightarrow \mathrm{Y} \rightarrow \mathrm{Z} \rightarrow \mathrm{X}$ occurs within the causal Bayesian network. The DAG structure avoids paradoxical situations where X causes itself since X cannot be its own cause. Finally, the L1 regularization term of the causal Bayesian network parameters, denoted as $\mathcal{L}_{\text{s}}$, is used to enforce sparsity in the graph, making it easier to interpret the causal graph.

3.2 Causal Deep Learning

Deep learning models typically generate predictions by identifying associations between feature X and label Y, i.e., likelihood $P(Y|X)$. However, when X and Y have a common ancestor Z (confounding variable, as shown in Fig. 2a), learning $P(Y|X)$ will lead the model to learn spurious relationships [20]. Because

$$P(Y|X) = \sum_{z} P(Y|X,z)\,P(z|X), \tag{2}$$

the confounder Z introduces the bias through $P(z|X)$. For instance, suppose that in the repost prediction scenario, P (z = public media | X = anime) is greater than P (z = friends | X = anime). In that case, most of the likelihood in Eq. (2) will arise from P (Y = repost | X = anime, z = public media) instead of P (Y = repost | X = anime, z = friends). When predicting a user's repost behavior for posts about anime, the model tends to prioritize posts shared by the public media over the content of the post itself.

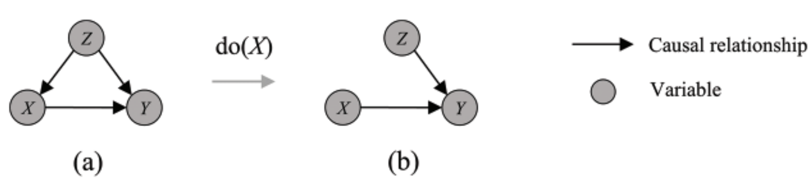

Figure 2: The diagram of do-calculus

Therefore, recent deep learning models [11,19] implement causal learning by leveraging the do-calculus to block the causal relationship between Z and X, as shown in Fig. 2b. These models make predictions based on the causal association $P(Y|\text{do}(X))$ between X and Y, where

$$P(Y|\text{do}(X)) = \sum_{z} P(Y|X,z)\,P(z). \tag{3}$$

Compared to Eq. (2), z is no longer affected by X, and thus the do-calculus deliberately forces X to incorporate every z fairly, subject to its prior $P(z)$, into the prediction of Y.

In the context of the repost prediction, addressing the impact of confounding variables is equally crucial. Employing causal deep learning techniques enables these models to discern causal associations and enhances their generalization capabilities. However, two substantial challenges are encountered:

- To perform causal deep learning (as illustrated in Fig. 2), a model needs to initially identify the confounding variable Z. However, in the context of repost prediction involving multiple variables such as social relationships (S), user interactions (D), user topic interests (I), and query posts (T), obtaining complete prior knowledge of the causal relationships between these variables poses a significant challenge. Depending solely on expert knowledge makes it difficult to ascertain the causal relationships between variables. For instance, in tasks such as Visual Question Answering (VQA), diverse experts may hold varying interpretations of the causal links between variables [11,13].
- As depicted in Eq. (3), when a model learns causal association, it needs to know the distribution of the confounder Z. By leveraging prior knowledge of causality, it becomes possible to amass a large number of instances related to Z before model training. This facilitates the creation of a confounder dictionary, serving as a representation of the distribution of Z [11,13,26]. In the absence of prior knowledge of causality, the construction of confounder dictionaries becomes unfeasible. This impedes the ability of the repost prediction model to discern the causal association among variables.

4 Method

4.1 Problem Formulation and Notations

The repost prediction task can be seen as a classification task [5]. Given a query post T and a user u, the repost prediction task is to predict whether u will repost based on T and the microblogging data. These data include (1) past posts reposted by users $I = \{t_1, \ldots, t_l\}$; (2) the following network $F = \{U_F, E_F\}$, where U_F is the set of users and E_F is the set of following relationships of users; (3) the temporal repost interaction network $D = \{D_1, \ldots, D_d\}$, $D_m = \langle U_m, E_m, d_m \rangle$, where U_m and E_m are the sets of users and interaction relations in time slice d_m, $1 \leq \mathrm{m} \leq \mathrm{d}$, respectively. To improve the generalization ability of the repost prediction model, we combine it with causal inference to control the impact of confounding variables to achieve Debiased Repost Prediction (DRP). However, as it is difficult for us to have complete prior knowledge of the causal relationships of variables, we introduce Causal Discovery methods into DRP model, i.e., CD-DRP model (see Fig. 3). The proposed CD-DRP model can achieve repost prediction by observing microblogging data without making assumptions about the causal relationships between variables.

The CD-DRP model diverges fundamentally from existing causal deep learning models, such as those presented in [11,19]. Firstly, the CD-DRP model operates independently of expert knowledge on causal relationships, autonomously discovering variable causality through the observation of microblogging data. This autonomy allows the model to adeptly handle repost prediction scenarios involving multiple variables and mitigate bias stemming from expert knowledge. Secondly, in the process of learning causal associations, the CD-DRP model obviates the need for manually constructing confounder dictionaries. This characteristic provides effective support for debiased repost prediction.

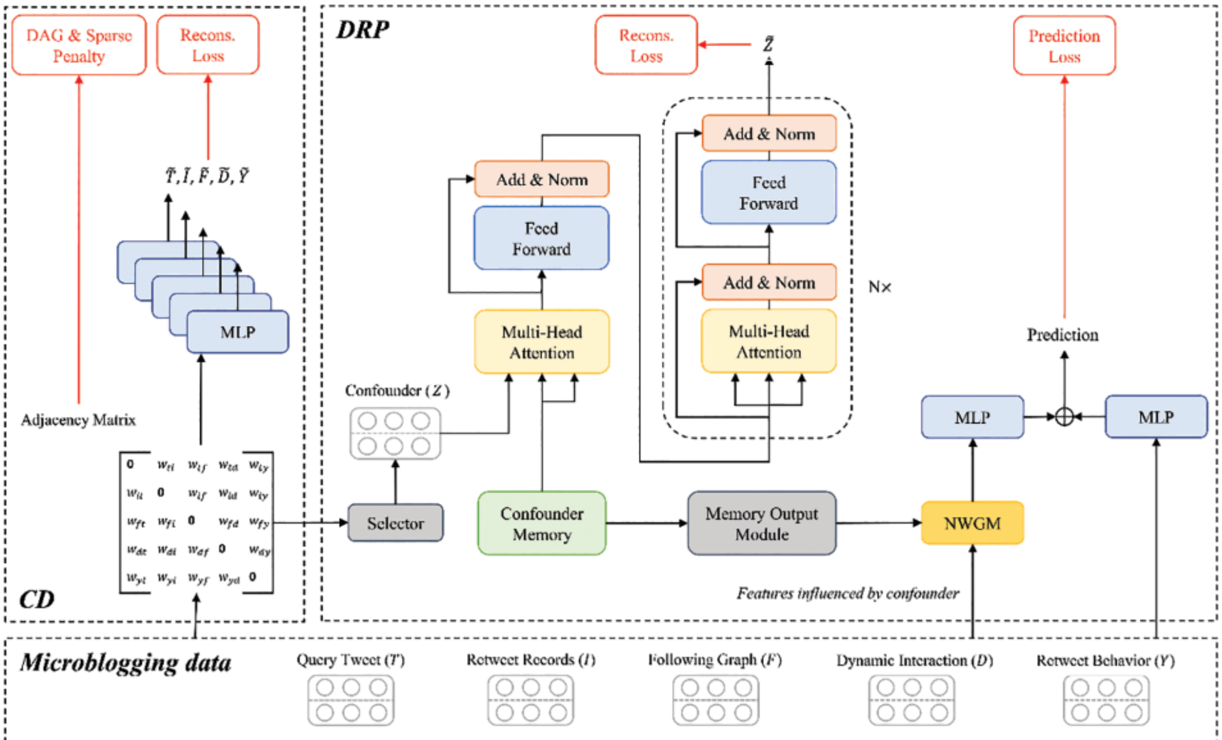

Figure 3: The framework of proposed CD-DRP

4.2 Embedding

To facilitate the proposed CD-DRP model to learn microblogging data, we first map T, I, F, and D to a multidimensional feature space. Specifically, T and I composed of text use the pre-trained language model BERT [30] as the embedding method. If a query post is t_q, its feature representation $x_t = \text{BERT}\left(t_q\right)$, $x_t \in R^{d_t}$. d_t is the dimension of the BERT feature. If the repost record of user u is $I_u = \{t_1^u, \ldots, t_l^u\}$, the feature of I_u is represented as $X_u = \{x_1^u, \ldots, x_l^u\}$, $x_j^u = \text{BERT}\left(t_j^u\right)$, $1 \le j \le l$, $x_j^u \in R^{d_t}$. We perform an average pooling operation on X_u to obtain the feature representation x_i of the user interests, $x_i = \text{AvgPooling}\left(X_u\right)$, $x_i \in R^{d_t}$. For the graph structure data F and D, we use node2vec [31] as the embedding method. The feature of F is represented as $X_F = \text{node2vec}\left(F\right)$, $X_F \in R^{n \times d_g}$, n is the number of users, and d_g is the node2vec feature dimension. Each row of X_F represents the features x_u of a user in the following network. Each time slice of repost interaction network D is also embedded with node2vec, $X_D = \{X_{D_1}, \ldots, X_{D_d}\}$, $X_{D_m} = \text{node2vec}\left(D_m\right)$, $X_D \in R^{d \times n_m \times d_g}$, and n_m is the number of users of time slice m. Take the features of user u in different time slices of the repost interaction network and do the averaging pooling operation to get the representation x_d of the user interaction preference, $x_d = \text{AvgPooling}\left(x_D^u\right)$, $x_d \in R^{dg}$, $x_D^u = \{x_{D_1}^u, \ldots, x_{D_d}^u\}$.

4.3 Causal Discovery

The aim of the causal discovery (CD) module is to identify the causal graph with the most explanatory power for the microblogging data and assist the debiased repost prediction module in identifying confounding variables. As variables like T, I, F, and D are multidimensional, discovering the causal graph through a single linear model or MLP, as in recent work [27,28], is not feasible.

Therefore, based on the paradigm described in Section 3.1, we develop a deep generative network that can uncover the causal graph for multidimensional feature variables. This network can gradually minimize the difference between the observed and conditional expectation $E(X|\text{Parent}(X))$ of the multidimensional variable X by modifying its parameters. Specifically, the causal discovery module comprises a learnable parameter matrix (causal graph) $\mathcal{A}$ and multiple parallel MLPs for generation (as illustrated in Fig. 3), where

$$\mathcal{A} = \begin{bmatrix} 0 & w_{ti} & w_{tf} & w_{td} & w_{ty} \\ w_{it} & 0 & w_{if} & w_{id} & w_{iy} \\ w_{ft} & w_{fi} & 0 & w_{fd} & w_{fy} \\ w_{dt} & w_{di} & w_{df} & 0 & w_{dy} \\ w_{yt} & w_{yi} & w_{yf} & w_{yd} & 0 \end{bmatrix}, \tag{4}$$

the parameters in the columns of matrix $\mathcal{A}$ represent the weight of the incoming edges or the influence of other variables on a given variable. For instance, the first column of $\mathcal{A}$, denoted as $\mathcal{A}_{:\mathcal{T}} = \left[0, w_{it}, w_{ft}, w_{dt}, w_{yt}\right]$, represents the edges pointing to node T from nodes T, I, F, D, and Y. The main diagonal elements of $\mathcal{A}$ are fixed at 0 to indicate that each variable cannot be self-explained. The parental information of any variable X expresses as $\text{Parent}(X) = \mathcal{A}_{:\mathcal{X}}^{T} \cdot \mathcal{M}$, where $\mathcal{M} = [T, I, F, D, Y]$. Given the parental information, the conditional expectation of X is

$$\begin{aligned} \tilde{X} &= E(X|\text{Parent}(X)) \\ &= \text{MLP}\left(A_{:X}^{T} \cdot \mathcal{M}\right) = \sigma\left(W^{(h)}\sigma\left(\ldots\sigma\left(W^{(1)}A_{:X}^{T} \cdot \mathcal{M}\right)\right)\right), \end{aligned} \tag{5}$$

where σ is the activation function, e.g., ReLU, and $W^{(k)}$ is the parameter of the k-th perceptron layer, $1 \le k \le h$. According to Eq. (5), we can obtain the conditional expectations of different variables from a number of MLPs. The generated microblogging data can be represented as $\tilde{\mathcal{M}} = \left[\tilde{T}, \tilde{I}, \tilde{F}, \tilde{D}, \tilde{Y}\right]$.

During training, the causal discovery module continually optimizes the parameters of both $\mathcal{A}$ and MLPs to minimize the reconstruction loss of the microblogging data,

$$\mathcal{L}_{rec} = \text{RMSE}\left(\mathcal{M}, \tilde{\mathcal{M}}\right), \tag{6}$$

ultimately achieves the goal of uncovering the causal relationships between variables and understanding the mechanisms that drive the generation of data. Here, RMSE is the root mean square error, which is used to measure the discrepancy between $\mathcal{M}$ and $\tilde{\mathcal{M}}$. In addition, in order to ensure $\mathcal{A}$ is a sparse DAG, the causal discovery module also imposes the following constraints:

$$\mathcal{L}_{DAG} = \text{tr}\left(e^{\mathcal{A} \cdot \mathcal{A}}\right) - d, \tag{7}$$

$$\mathcal{L}_{s} = |\mathcal{A}|_1, \tag{8}$$

where $\mathcal{L}_{DAG}$ is proposed by [27], $\text{tr}(\cdot)$ denotes the trace of the matrix, $e^{\mathcal{A} \cdot \mathcal{A}}$ is the matrix exponential of Hadamard product of $\mathcal{A}$, d is the number of elements of the main diagonal of $\mathcal{A}$, and $|\cdot|_1$ is the L1 parametrization. Therefore, the objective of the causal discovery module is

$$\begin{aligned} \mathcal{L}_{CD} &= \mathcal{L}_{rec} + \alpha\mathcal{L}_{DAG} + \beta\mathcal{L}_{s} \\ &= \text{RMSE}\left(\mathcal{M}, \tilde{\mathcal{M}}\right) + \alpha\left(\text{tr}\left(e^{\mathcal{A} \cdot \mathcal{A}}\right) - d\right) + \beta|\mathcal{A}|_1. \end{aligned} \tag{9}$$

4.4 Debiased Repost Prediction

Based on the discovered causal graph, we can identify confounders that affect the generalizability of the repost prediction model, i.e., variables that affect both predictor X and label Y such as Z in Fig. 2a. To control this influence, we devise the Debiased Repost Prediction (DRP) module, which introduces causal inference to the repost prediction model as mentioned in Section 3.2. For the predictor X affected by confounders, given its feature x and user repost behavior $Y = y$, we can write the causal association between X and Y as $\sum_z P(y|x,z)\,P(z)$. Here, the conditional probability $P(y|x,z)$ can be parameterized by a Softmax layer, i.e., $P(y|x,z) = \mathrm{Softmax}(f(x,z))$, where $f(\cdot)$ is a function that calculates the probability of various user behaviors. $\sum_z \cdot P(z)$ is the expectation based on the prior distribution of z. Therefore, we implement DRP as

$$P(Y|\mathrm{do}(X)) = \mathbb{E}_z[\mathrm{Softmax}(f(x,z))]. \tag{10}$$

However, $\mathbb{E}_z$ requires expensive sampling, which can significantly affect the training efficiency. Therefore, we adopt the scheme of [11,26] and use the Normalized Weighted Geometric Mean (NWGM) [32] to approximate this expectation, i.e.,

$$\mathbb{E}_z[\mathrm{Softmax}(f(x,z))] \approx \mathrm{Softmax}(\mathbb{E}_z[f(x,z)]). \tag{11}$$

By setting $f(\cdot)$ to a linear model, $f(x,z) = W_x x + W_z z$, Eq. (10) can be written as:

$$P(Y|\mathrm{do}(X)) = \mathrm{Softmax}(W_x x + W_z \cdot \mathbb{E}_z(z)), \tag{12}$$

where W_x and W_z are linear neural network parameters. As soon as $\mathbb{E}_z(z)$ is obtained, we can evaluate the causal association between X and Y according to Eq. (12). For a predictor X_n that is not affected by confounding variables, given $X_n = x_n$, the causal association between X_n and Y is $P(y|x_n)$, implemented as:

$$P(Y|X_n) = \mathrm{Softmax}(W_n x_n), \tag{13}$$

with W_n is the parameter of the linear neural network. Based on the causal association of different predictors with user behavior Y, the DRP module makes the prediction (as shown in Fig. 3):

$$\tilde{y} = \mathrm{Softmax}(W_x x + W_z \cdot \mathbb{E}_z(z)) + \mathrm{Softmax}(W_n x_n). \tag{14}$$

During training, the DRP module takes the cross entropy $\mathcal{L}_p$ between the predicted value $\tilde{y}$ and the label y as the objective function,

$$\mathcal{L}_p = -\sum_y y \log(\tilde{y}), \tag{15}$$

to optimize the parameters and improve its prediction performance.

It should be noted that CD-DRP lacks prior knowledge of the causal relationships and therefore cannot identify Z before training to construct a confounder dictionary that represents $P(z)$, as demonstrated in recent studies [11,13,26]. Instead, CD-DRP needs to estimate $P(z)$ through training, utilizing information on confounders that are scattered across different data batches. It presents a challenge in computing $\mathbb{E}_z(z)$. To overcome this issue, we devise a novel confounder memory network. This network is parameterized as a matrix $\mathcal{C}$, $\mathcal{C} \in R^{s\times dc}$. While training, the DRP module retrieves $\mathcal{C}$'s memory $\tilde{z}$ for different data batches of confounder z, and continuously reduces the reconstruction loss of $\tilde{z}$ and z to optimize $\mathcal{C}$'s parameters. In this way, $\mathcal{C}$ is able to perceive $P(z)$. Specifically, when retrieving $\mathcal{C}$'s memory of z, the DRP module first calculates the association of z to each piece of information stored by $\mathcal{C}$:

$$\alpha_z = \text{Softmax}\left(z^T\mathcal{C}/\sqrt{d_c}\right), \tag{16}$$

the information about z is $z' = \alpha_z\mathcal{C}$. Subsequently, DRP employs multiple self-attention [33] layers to represent the memory $\tilde{z}$.

$$\tilde{z} = \text{Self}-\text{attention}^{(N)}\left(\ldots\text{Self-attention}^{(1)}\left(\alpha_z\mathcal{C}\right)\right). \tag{17}$$

The reconstruction loss of z is the root mean square error:

$$\mathcal{L}_{zrec} = \text{RMSE}\left(z, \tilde{z}\right). \tag{18}$$

After multiple rounds of training, $\mathcal{C}$ progressively saves information on confounding variables in different data batches. Finally, DRP performs an average pooling operation on $\mathcal{C}$ to calculate $\mathbb{E}_z\left(z\right)$:

$$\mathbb{E}_z\left(z\right) = \text{AvgPooling}\left(\mathcal{C}\right). \tag{19}$$

To summarize, the DRP module is designed to achieve two objectives: to perceive the distribution of confounding variables and to predict user behavior. The objective for the module is denoted as $\mathcal{L}_{DRP}$.

$$\begin{aligned}\mathcal{L}_{DRP} &= \mathcal{L}_p + \mathcal{L}_{zrec} \\ &= -\sum_y y\log\left(\tilde{y}\right) + \text{RMSE}\left(z, \tilde{z}\right).\end{aligned} \tag{20}$$

4.5 CD-DRP

The descriptions of different modules of the proposed CD-DRP model in Sections 4.2–4.4 are summarized here as Algorithm 1. Specifically, the CD-DRP model first uses BERT [30] and node2vec [31] to map query posts T, user repost records I, following network F, and repost interaction network D to a multidimensional feature space (Section 4.2). Then, the CD-DRP model identifies the parents of different variables based on the causal graph $\mathcal{A}$ and generates different variables according to Eq. (5). By optimizing Eq. (9), the CD-DRP model is able to update the parameters of the causal discovery module and find the causal graph $\mathcal{A}$ with the best explanatory power for the microblogging data (Section 4.3). Finally, the CD-DRP model identifies the confounder Z with the help of $\mathcal{A}$. The distribution of Z in the training set is perceived by Eqs. (16) and (17). By optimizing the loss calculated by Eq. (20), CD-DRP is able to make prediction with the help of Eq. (14) (Section 4.4).

Algorithm 1: Causal Discovery for Debiased Repost Prediction (CD-DRP)

Input: The query post T, user repost records I, following network F, repost interaction network D, and user repost behavior Y.
Output: Parameters Θ of the CD-DRP, $\Theta = \{\Theta_{CD}, \Theta_{DRP}\}$.
1: Embedding T, I, F, and D using BERT and node2vec.
2: Initialize Θ.
3: **for** *every epoch* **do**
4: **for** *every batch* **do**
5: identify the parents of T, I, F, D, and Y via causal graph $\mathcal{A}$.
6: generate conditional expectations for different variables via Eq. (5).
7: evaluate $\mathcal{L}_{CD}$ via Eq. (9).
8: update Θ_{CD} according to the gradient decent of $\mathcal{L}_{CD}$.

(Continued)

Algorithm 1 (continued)

```
    end for
    for every batch do
        identify confounders via 𝒜.
        retrieve the confounder z̃ via Eqs. (16) and (17).
        predict user behavior ỹ via Eq. (14).
        evaluate ℒ_DRP via Eq. (20).
        update Θ_DRP according to the gradient decent of ℒ_DRP.
    end for
end for
Return Parameters Θ of CD-DRP.
```

5 Experiments

In this section, we conduct experiments to evaluate the effectiveness of the proposed model with a comparison to state-of-the-art baselines. Specifically, we aim to answer the following research questions:

- **RQ1:** How does the generalizability of the proposed CD-DRP model compare to the state-of-the-art baselines?
- **RQ2:** What causal relationships are identified by the CD-DRP model? Do these relationships align with common perceptions?
- **RQ3:** If CD-DRP model performs well, what component benefits CD-DRP model in the repost prediction task?
- **RQ4:** What is the influence of different hyperparameter settings on the performance of CD-DRP model?
- **RQ5:** Compared to baselines, what is the time efficiency of CD-DRP model?

5.1 Experimental Settings

Datasets The proposed CD-DRP model is assessed through experiments using the Twitter-Dynamic-Net (TDN)[1] and Weibo-Net-Tweet (WNT)[2] datasets. To ensure result reliability, we draw three samples from each dataset for evaluation. The sampling procedure follows these steps: First, for both TDN and WNT, we randomly select 2000 users and gather their follower lists. Next, the repost records $\mathcal{R}_x = \{< \mathrm{t}_1, \mathrm{d}_1, \mathrm{u}_1, \mathrm{a}_1 >, \ldots, < \mathrm{t}_i, \mathrm{d}_i, \mathrm{u}_i, \mathrm{a}_i >, \ldots, < \mathrm{t}_n, \mathrm{d}_n, \mathrm{u}_n, \mathrm{a}_n >\}$ of these users are considered as positive instances, with the repost history being the reposts before d_i, where t_i refers to the post, d_i is the repost date, and u_i and a_i denote the user and post publisher, respectively. Posts that have not been reposted by users are randomly selected as negative instances, maintaining a 1:1 ratio with the positive instances. For each sample, the training set covers three months, while the test set consists of data from the following ten days. Finally, we extract dynamic interactions $\mathcal{I}_x = \{< d_1, u_1, a_1 >, \ldots, < d_i, u_i, a_i >, \ldots, < d_n, u_n, a_n >\}$ from the repost records, which are used to construct the dynamic interaction graph, ensuring that $\mathcal{I}_x$ only includes interactions from the training set to prevent test set leakage. Table 1 summarizes the details of the experimental datasets.

[1] https://www.aminer.cn/data-sna#Twitter-Dynamic-Net (accessed on 13 November 2024).
[2] https://www.aminer.cn/data-sna#Weibo-Net-Tweet (accessed on 13 November 2024).

Table 1: Details of the experimental datasets[1]

Dataset	Sample	Users	Posters	Pos.	Neg.	Histories
TDN	1	707	3168	8414	8414	10,044
	2	750	3293	8435	8425	9430
	3	767	3554	11,082	11,053	12,195
WNT	1	1443	7109	10,527	10,100	19,327
	2	1504	7594	10,816	10,464	20,166
	3	1454	7349	10,488	10,043	19,942

Note: [1]It should be noted that the number of users in each dataset sample is less than 2000 due to the possibility of randomly selected users not exhibiting repost behavior during the time period in which the dataset samples were collected.

Baselines We conduct a comparative analysis of the proposed method against various baseline models, encompassing random models, those with simple structures (such as LR, SUA-ACNN [34], GraphSAGE [35], and DynamicGCN [36]), and those with complex structures (including AMNL [37], DFMF [38], GFCI [39], and GCRec [15]):

- **Random:** The prediction of reposting behavior, whether it occurs or not, is randomly determined in this model, which establishes the distribution of positive and negative instances and serves as the baseline. Any model capable of learning from data to predict reposting behavior is expected to outperform this Random model.
- **LR:** A simple classification model. It predicts user behavior based on BERT features of the query post and repost records.
- **SUA-ACNN:** An attention-based convolutional neural network. This method uses convolutional neural networks to capture features of posts and learns user interests with attention mechanisms. By evaluating the similarity between query posts and users' interests, SUA-ACNN makes predictions about users' behavior.
- **GraphSAGE:** A graph convolutional network. It learns the topological information of following network with graph convolution and uses it to represent the social relationship features of users. The social relationship features are the basis for prediction.
- **DynamicGCN:** A dynamic graph neural network, which uses graph convolutional networks and recurrent neural networks to capture the features of temporal repost interactions. DynamicGCN predicts user behavior based on their past repost interactions.
- **AMNL:** This model constructs a heterogeneous network consisting of following relationships, repost interactions, and posts. It learns the joint post representations and user preference representations from this heterogeneous network for repost prediction.
- **DFMF:** This model uses deep representation learning methods to capture the following network and post features. Based on these multimodal features, it employs a fully-connected forward neural network to predict user behavior.
- **GFCI:** This model utilizes the bidirectional attention mechanism to capture user interest from past reposts and learns the features of repost interactions through a dynamic graph convolutional network. It makes the prediction through a multimodal fusion layer.
- **GCRec:** A causal-based reposting prediction model, which first utilizes graph neural networks to capture the information of the social and reposting interaction networks, representing the features of users and posts. Subsequently, do-calculus (see Section 3.2) is applied to control for

the influence of confounding variables, enabling the prediction of reposting behavior based on users' preferences. The variable causal diagram is identified by our CD-DRP model.

Metrics The experiments use the popular Accuracy (A) and F1-score (F) to quantify the repost prediction performance of the proposed model and baselines. The definition of A and F are as follows:

$$A = \frac{tp + tn}{tp + fn + tn + fp}, \tag{21}$$

$$F = \frac{2 \times P \times R}{P + R} \quad P = \frac{tp}{tp + fp} \quad R = \frac{tp}{tp + fn}, \tag{22}$$

where tp is the number of true positive, tn is the number of true negative, fp the number of false positive, and fn is the number of false negative.

5.2 Implementation Details

We implement the proposed model and all baseline models (excluding LR) using PyTorch, while LR is implemented with scikit-learn's built-in function. The pretrained BERT parameters are obtained from the links provided in the footnotes[3,4]. Both BERT and node2vec use an embedding dimension of 768, and the memory network size for CD-DRP is set to 512 × 768. CD-DRP also includes 6 self-attention layers. The Adam optimizer is used for training all models. For all models except LR, parameters are initialized with the Glorot method. The learning rate is set to 0.0001, L2 regularization weights are 0.001, the dropout rate is 0.5, and the batch size is 128. Early stopping is applied with a patience of 10 epochs, using F1-score as the metric to monitor. All experiments are conducted on an NVIDIA RTX 3080 GPU (10 GB) and an Intel i7-10700F CPU (64 GB).

5.3 Generalizability (RQ1)

Generalizability is important for repost prediction models because the goal of these models is to be able to accurately predict users' repost behavior for new, unseen posts. A model with poor generalizability may suffer from overfitting problems, i.e., it performs well on training data but fails to make accurate predictions on testing data. Therefore, we analyze the generalizability of the proposed CD-DRP model and baselines in terms of both prediction performance and overfitting degree.

Prediction Comparison Table 2 records the prediction performance of the proposed CD-DRP model and baselines on the testing set of six sample datasets. From Table 2, we can find that:

- Each of the simple baseline methods demonstrates superior performance compared to the Random model. This observation indicates that the inclusion of the query post, user repost records, following networks, and repost interactions as features is advantageous for predicting user behavior. Notably, GraphSAGE and DynamicGCN exhibit comparable performance, with a marginal difference in average test Accuracy of less than 0.7%. Moreover, both GraphSAGE and DynamicGCN outperform LR and SUA-ACNN in the majority of cases.
- The complex structural models exhibit a significant improvement over the simple structural models across multiple dataset samples. For instance, DFMF surpasses DynamicGCN in different dataset samples, resulting in an average improvement of 4.07% in test Accuracy and 2.46% in F1-score. Among the complex structural models, the causal-based GCRec model outperforms the others, while the remaining models show comparable performance.

[3]Parameters from https://huggingface.co/roberta-base for TDN (accessed on 13 November 2024)
[4]Parameters from https://huggingface.co/bert-base-chinese for WNT (accessed on 13 November 2024)

Specifically, the GCRec model achieves an average F1-score that is approximately 3.7% higher than that of the GFCI model across all datasets. In contrast, the GFCI model's performance is close to that of the AMNL and DFMF models, with an average F1-score difference of less than 2%. These results underscore the importance of controlling for confounding variables and predicting user reposting behavior based on causal associations. Since these models learn similar information—such as query posts, user reposting histories, following relationships, and reposting interaction networks—the GCRec model demonstrates superior ability to capture user reposting preferences.

- The prediction performance of the proposed CD-DRP model outperforms all baselines. For the state-of-the-art baseline GCRec, CD-DRP improves both Accuracy and F1-score metrics by about 2.90% and 2.54%. This result confirms the CD-DRP model's ability to control for confounding variables and enhance reposting prediction performance. Additionally, by discovering causal relationships through the analysis of online social network data, the CD-DRP model demonstrates greater practical value compared to other causal-based reposting prediction models.

Table 2: Prediction performance of CD-DRP and baselines

	Models	Sample 1		Sample 2		Sample 3	
		A	*F*	*A*	*F*	*A*	*F*
TDN	Random	0.495	0.499	0.483	0.488	0.493	0.494
	LR	0.545	0.559	0.525	0.536	0.561	0.574
	SUA-ACNN	0.659	0.697	0.622	0.549	0.631	0.560
	GraphSAGE	0.733	0.700	0.683	0.661	0.761	0.760
	DynamicGCN	0.736	0.725	0.702	0.708	0.727	0.755
	AMNL	0.752	0.771	0.710	0.737	0.775	0.796
	DFMF	0.756	0.763	0.723	0.722	0.787	0.790
	GFCI	0.766	0.761	0.734	0.716	0.807	0.798
	GCRec	0.775	0.783	0.726	0.747	0.776	0.801
	CD-DRP	**0.788**	**0.798**	**0.740**	**0.761**	**0.810**	**0.820**
WNT	Random	0.526	0.532	0.490	0.495	0.503	0.507
	LR	0.618	0.597	0.639	0.626	0.629	0.617
	SUA-ACNN	0.618	0.636	0.499	0.665	0.592	0.653
	GraphSAGE	0.638	0.637	0.622	0.628	0.643	0.637
	DynamicGCN	0.635	0.613	0.618	0.593	0.662	0.667
	AMNL	0.666	0.625	0.665	0.610	0.677	0.681
	DFMF	0.642	0.584	0.673	0.650	0.665	0.654
	GFCI	0.678	0.632	0.670	0.632	0.668	0.626
	GCRec	0.684	0.647	0.694	0.663	0.691	0.672
	CD-DRP	**0.701**	**0.668**	**0.718**	**0.689**	**0.715**	**0.685**

Overfitting Comparison As described in Section 5.2, we utilize the F1-score as an early stopping metric during training and assess the model's repost prediction ability on the training set. To evaluate the degree of overfitting, we examine the difference, denoted as ΔF, between the F1-score of the model on the training and testing sets. A smaller value of ΔF indicates a closer alignment between the model's prediction performance on the training and testing sets, thereby suggesting a lower degree of overfitting. Since generalizability encompasses multiple aspects, we aim for a balance between predictive performance and overfitting. Ideally, a model with good generalizability should exhibit strong prediction performance while minimizing overfitting.

In this study, we compare the degree of overfitting of the proposed CD-DRP model and the 4 best baselines in terms of prediction performance (shown in Fig. 4). As can be seen from Fig. 4, these models have an F1-score distributed between 0.96 and 1.00 at the end of the training. They all fit the training data well. When combined with their F1-score on the testing set recorded in Table 2, we can see that CD-DRP model has the smallest ΔF on most of the dataset samples (except WNT Sample 3). Specifically, the overfitting of CD-DRP model was reduced by 7.44%, 9.18%, 11.96%, and 16.84% compared to GCRec, AMNL, DFMF, and GFCI, respectively. This result shows that the proposed CD-DRP model combined with causal discovery and inference can provide a more realistic understanding of microblogging data and alleviate the overfitting problem.

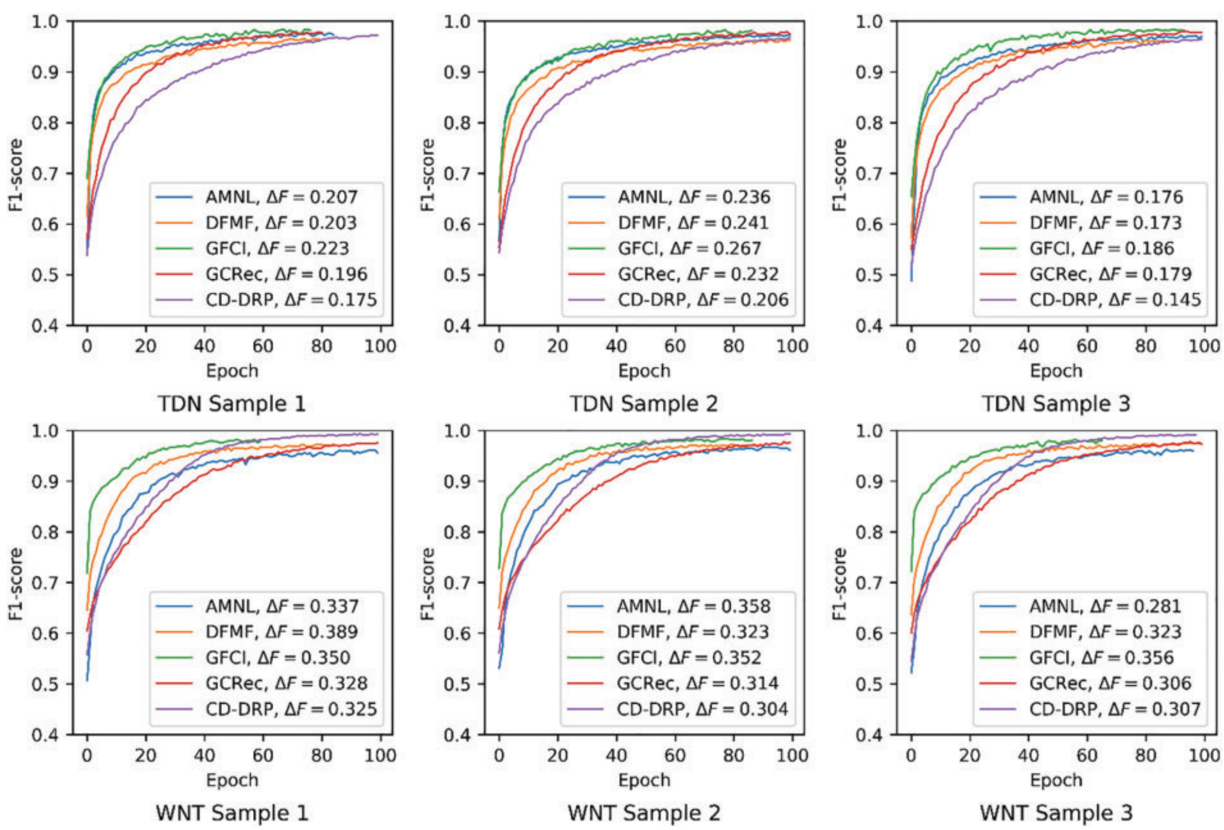

Figure 4: The F1-score of CD-DRP model and comparison models on the training set as the training progresses. ΔF is the difference between the F1-score of the model on the training set and the testing set at the end of training

5.4 Causal Graphs (RQ2)

This section presents the causal graphs identified by CD-DRP model in Fig. 5. It can be observed that query posts T, user interests I, following relations F, and interaction relations D are the parents of the repost behavior Y, implying that these variables contribute to predicting repost behavior. This finding is reinforced by the results presented in Table 3. For instance, the simple structural model DynamicGCN, which uses information from D, outperforms Random across various datasets. Meanwhile, the following relationships F serve as the parent node for both query posts T and user behaviors Y. This means that the posts a user sees and his repost behavior are influenced by the following relationship, which aligns with previous research findings [40,41]. However, we also observe a discrepancy in the causal graphs identified between the TDN and WNT datasets. Specifically, in the TDN dataset, the variable I serves as the parent node of D ($I \rightarrow D$), while in the WNT dataset, I functions as the parent node of T ($I \rightarrow T$). This phenomenon might be influenced by disparities between the X and Sina Weibo platforms. For example, Li et al. [42] found that users on platform X tend to engage in interactions with sports organizations, whereas Sina Weibo users prefer collecting information and expressing support for sports teams.

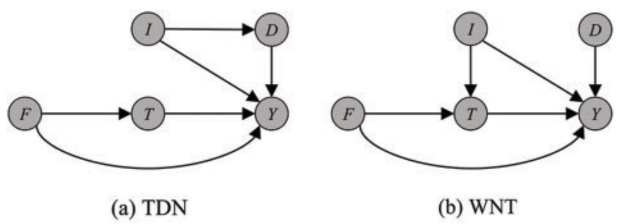

Figure 5: The identified causal graphs for repost prediction scenarios

Table 3: The influence of modifying the $I \rightarrow D$ and $I \rightarrow T$ relationships on the CD-DRP model

	Models	Sample 1		Sample 2		Sample 3	
		A	*F*	*A*	*F*	*A*	*F*
TDN	CD-DRP^{m}	0.753	0.772	0.735	0.748	0.790	0.801
	CD-DRP	**0.788**	**0.798**	**0.740**	**0.761**	**0.810**	**0.820**
WNT	CD-DRP^{m}	0.675	0.629	0.711	0.674	0.709	0.676
	CD-DRP	**0.701**	**0.668**	**0.718**	**0.689**	**0.715**	**0.685**

To validate the causal relationships that exhibit differences between the TDN and WNT datasets, this section further conducts a causal analysis of these variations. We employ the method adopted by [43,44] to validate the causal relationships $I \rightarrow D$ depicted in Fig. 5a and $I \rightarrow T$ illustrated in Fig. 5b. Specifically, we leverage *Permutation Importance* (PI) for causal validation. The directions of causal relationships $I \rightarrow D$ and $I \rightarrow T$ are permuted. We train the CD-DRP model to capture variable associations by adapting to the modified causal graphs. The performance of the updated model CD-DRP^{m} is then evaluated in terms of prediction Accuracy and F1-score. A decrease in model prediction performance following the modification suggests the correctness of the original $I \rightarrow D$ and $I \rightarrow T$ relationships. Conversely, a performance improvement implies that the initial $I \rightarrow D$ and $I \rightarrow T$ relationships were incorrect. Table 3 provides a clear comparison of performance metrics, indicating that the CD-DRP^{m} model consistently exhibits lower performance across all sampled datasets when compared to the CD-DRP model. This observation underscores the significance of the causal relationship where I serves as the parent node of D in the TDN dataset and as the parent node of T in the WNT dataset. Understanding these causal connections is crucial for enhancing the model's comprehension of microblogging data, leading to improved predictions of user retweet behavior.

5.5 Ablation Study (RQ3)

As detailed in Section 1, our primary objective is to safeguard the repost prediction model from being influenced by confounding factors, which could otherwise result in the learning of spurious relationships between features and user repost behaviors. To achieve this, we introduce causal discovery and inference methods into the repost prediction model, referred to as CD-DRP. The CD-DRP model comprises two key modules: a causal discovery module responsible for identifying causal relationships among variables, and a prediction module designed to mitigate the impact of confounding variables. Additionally, we incorporate a confounder memory network to enhance CD-DRP's ability to perceive the distribution of confounders.

In this study, we investigate the effects of these modules on the performance of CD-DRP model through an ablation study. Firstly, we randomly disrupt the causal relationships identified by the causal discovery module and employ the disrupted relationships to guide the prediction module in generating debiased repost predictions. This variant is denoted as CD-DRP (w/o Causal Discovery). Secondly, we omit the randomization of the causal discovery module's output, but refrain from making debiased predictions. Instead, we directly input different features into Multi-Layer Perceptrons (MLPs) for prediction. This model is labeled as CD-DRP (w/o Debiased Prediction). Lastly, we eliminate the confounder memory network and simply employ the mean of the pretrained features of confounders for debiased repost prediction. This variant is referred to as CD-DRP (w/o Memory Network).

The outcomes of the ablation study are presented in Fig. 6. From the results, it becomes evident that CD-DRP (w/o Causal Discovery) exhibits the poorest prediction performance. In terms of F1-score, it performs 4.04% lower than CD-DRP and even 0.28% lower than CD-DRP (w/o Debiased Prediction), which does not account for the impact of confounding variables. These findings highlight the crucial role of the causal relationships identified by the causal discovery module in guiding the prediction module to identify confounding variables and improve the generalization ability of CD-DRP model. Furthermore, by comparing the performance of CD-DRP (w/o Debiased Prediction) and CD-DRP, it becomes evident that controlling the impact of confounding variables is essential for CD-DRP. Without such control, the F1-score of CD-DRP experiences an average decrease of 3.74% across the six datasets examined. Lastly, upon removing the confounder memory network, the Accuracy and F1-score of CD-DRP demonstrate varying degrees of reduction across different datasets, with the exception of TDN Sample 2. On average, CD-DRP (w/o Memory Network) exhibits 1.69% lower Accuracy and 1.90% lower F1-score compared to CD-DRP model. These findings emphasize the significance of the confounder memory network in assisting CD-DRP model in better perceiving the distribution of confounding variables.

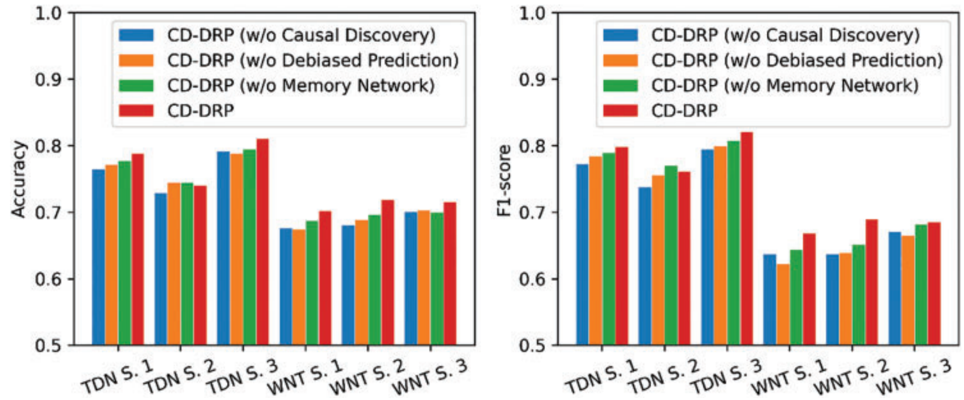

Figure 6: Performance comparison of different variants of the proposed CD-DRP

5.6 Hyperparameter Sensitivity (RQ4)

The CD-DRP model contains some unlearnable hyperparameters. We choose the following aspects to analyze the influence of hyperparameter settings on its prediction performance: (1) the learning rate, (2) the number N of self-attention layers of the confounder memory network, and (3) the size of the confounder memory network. We change one parameter in turn and fix other parameters to test the prediction performance of CD-DRP model in TDN Sample 2 and WNT Sample 2, the experimental results are shown in Fig. 7. From Fig. 7, it can be seen that the learning rate has a large influence on CD-DRP model. It is suitable for CD-DRP model to be trained at a small learning rate. Meanwhile, combined with the results on TDN and WNT, we find that CD-DRP model is not sensitive to the number of self-attention layers of its memory network. This indicates that if the application requires higher time efficiency, we can appropriately reduce the number of self-attention layers to achieve the goal of improving the model efficiency without affecting the model performance. Finally, it can also be seen from the results that CD-DRP model with a memory network size setting of 768×768 performs slightly worse than other settings. Therefore, the memory network size of CD-DRP model should be selected as 512×768 or smaller.

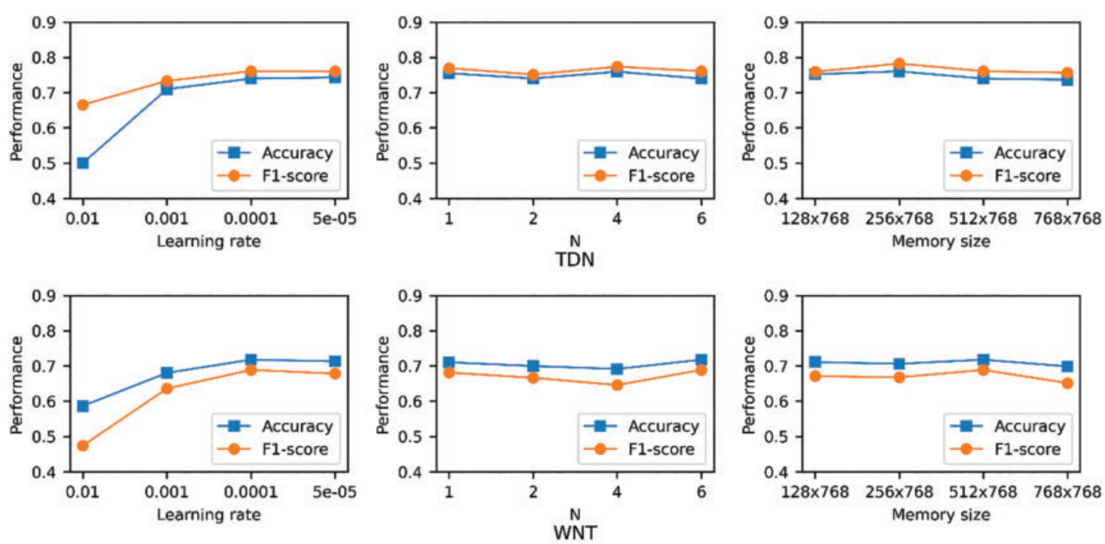

Figure 7: Prediction performance of CD-DRP under different hyperparameter settings

5.7 Time Efficiency (RQ5)

The experiments also analyze the time efficiency of different models to provide a basis for comparison and selection for practical applications. Specifically, we test the time it takes for different models to complete a batch training on TDN Sample 2 (as shown in Table 4). Based on the records in Table 4, it can be seen that DFMF has the best time efficiency, followed by GraphSAGE, and CD-DRP$_1$ and AMNL are close to and at a good level. Therefore, combined with their prediction performance (see Table 2), we can give preference to DFMF in application scenarios that require high time efficiency. In the scenario that needs to balance time efficiency and prediction power, we can choose the CD-DRP with a simplified structure (reduce the number of self-attention layers), such as CD-DRP$_1$.

Table 4: The time (ms) required for different models to complete a batch training, b_s is the batch size, and the subscript of CD-DRP is the number of self-attention layers used by its confounder memory network

Models	$b_s = 16$	$b_s = 32$	$b_s = 64$	$b_s = 128$	$b_s = 256$	$b_s = 512$
SUA-ACNN	39.451	40.433	35.320	39.237	47.562	82.414
GraphSAGE	7.876	8.858	12.329	15.649	24.603	43.779
DynamicGCN	54.884	58.859	60.597	62.469	62.591	75.387
AMNL	19.254	23.378	28.498	31.937	45.047	58.190
DFMF	**3.454**	**4.313**	**5.622**	**8.284**	**13.255**	**23.934**
GFCI	143.482	167.209	165.517	165.853	179.686	208.701
GCRec	19.688	32.612	61.700	105.665	202.020	391.938
CD-DRP$_1$	18.791	21.239	26.881	30.498	38.106	53.782
CD-DRP$_6$	42.259	46.332	54.850	64.950	69.800	83.431

5.8 Discussion

The above multi-dimensional quantitative analysis confirms that the CD-DRP model proposed in this paper demonstrates strong predictive performance and reduced overfitting. For instance, on the TDN sample dataset 2 (see Table 5), the CD-DRP model achieves a 1.87% improvement over the top-performing baseline, GCRec. Additionally, in terms of overfitting (measured by the F1-score difference between the end of training and testing), the CD-DRP model shows an 11.21% lower overfitting rate compared to the GCRec model. However, experimental results also reveal a gap between the CD-DRP model's testing F1-score and its F1-score at the end of training, indicating some degree of overfitting. Several factors may contribute to this: (1) The CD-DRP model's causal discovery paradigm (see Section 3.1) is data-driven, aiming to align the variable generation and observational distributions to uncover the causal structure most likely to represent the underlying variable generation mechanism. This may mean that the discovered causal diagram reflects only part of the true causal structure. (2) Test data distribution may differ from that of the training data. Since repost prediction requires a temporal component, training data is based on user activity from the past, while test data is drawn from later periods. Over time, online social network services can undergo changes influenced by various internal and external factors.

Table 5: Comparison of the CD-DRP model with state-of-the-art models

Model	Author	Year	F	ΔF
DFMF [38]	Yin et al.	2021	0.722	0.241
GFCI [39]	Sun et al.	2021	0.716	0.267
GCRec [15]	Yu et al.	2023	0.747	0.232
CD-DRP	Current study	2024	0.761	0.206

In prediction tasks using causal inference methods, confounding variable modeling typically relies on prior expert knowledge, with existing studies [11,15,26] identifying confounding variables by establishing a dictionary of these variables or pre-calculating their distributions before model training. However, in this study's context, these solutions are challenging to implement due to the limited prior knowledge of relevant causal relationships. To address this, we design a confounder memory network to simulate pre-representation of confounder features. This network identifies confounding variables during causal inference and dynamically updates its memory of these variables across data batches. Given its flexibility, this network is also suitable for other data mining tasks, supporting prediction tasks where causal relationships are not pre-defined.

6 Conclusion

Accurately predicting users' repost behavior is essential for opinion analysis and recommendation systems. Most existing models predict repost behavior by identifying associations between features and outcomes. However, these models often overlook confounding variables, which can lead to the learning of spurious relationships between features and user behavior, ultimately hindering their generalization ability. To address this issue, we propose CD-DRP, a model that performs causal discovery and DRP simultaneously. The proposed causal discovery module and confounder memory network enable us to control the influence of confounding variables, even in the absence of complete prior knowledge of variable causality. Experimental results demonstrate that the CD-DRP model surpasses the state-of-the-art model in terms of both prediction performance (with a notable improvement of 2.54%) and mitigating overfitting (with a reduction of 7.44%). In the future, our aim is to give the CD-DRP model the ability to mitigate the influence of hidden confounders, as collecting information on various aspects of microblogging services can be challenging. By doing so, we can further enhance the generalizability of the CD-DRP model and augment its value in real-world settings.

Acknowledgement: None.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Wu-Jiu Sun authored the main manuscript and conducted the experiments. Xiao Fan Liu contributed the theoretical concepts and made revisions to the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: In this study, we used a public dataset, which can be downloaded from the website if needed (https://www.aminer.cn/data-sna, accessed on 06 November 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] H. Luo, X. Meng, Y. Zhao, and M. Cai, "Exploring the impact of sentiment on multi-dimensional information dissemination using COVID-19 data in China," *Comput. Hum. Behav.*, vol. 144, no. 2, 2023, Art. no. 107733. doi: 10.1016/j.chb.2023.107733.

[2] D. MacKenzie, K. Caliskan, and C. Rommerskirchen, "The longest second: Header bidding and the material politics of online advertising," *Econ. Soc.*, vol. 52, no. 3, pp. 554–578, 2023. doi: 10.1080/03085147.2023.2238463.

[3] M. Ameri, E. Honka, and Y. Xie, "From strangers to friends: Tie formations and online activities in an evolving social network," *J. Mark. Res.*, vol. 60, no. 2, pp. 329–354, 2023. doi: 10.1177/00222437221107900.

[4] L. Guan, H. Liang, and J. J. H. Zhu, "Predicting reposting latency of news content in social media: A focus on issue attention, temporal usage pattern, and information redundancy," *Comput. Hum. Behav.*, vol. 127, no. 3, Mar. 2022, Art. no. 107080. doi: 10.1016/j.chb.2021.107080.

[5] J. Wang and Y. Yang, "Tweet retweet prediction based on deep multitask learning," *Neural Process Lett.*, vol. 55, no. 1, pp. 1–14, 2022. doi: 10.1007/s11063-021-10642-3.

[6] Y. Zhang and T. Hara, "Joint knowledge graph approach for event participant prediction with social media retweeting," *Knowl. Inf. Syst.*, vol. 66, no. 3, pp. 2115–2133, Sep. 2024. doi: 10.1007/s10115-023-02015-0.

[7] T. Zhong, J. Zhang, Z. Cheng, F. Zhou, and X. Chen, "Information diffusion prediction via cascade-retrieved in-context learning," in *Proc. 47th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, Washington, DC, USA, 2024, pp. 2472–2476. doi: 10.1145/3626772.3657909.

[8] F. F. Leung, F. F. Gu, Y. Li, J. Z. Zhang, and R. W. Palmatier, "Influencer marketing effectiveness," *J. Market.*, vol. 86, no. 6, pp. 93–115, Jun. 2022. doi: 10.1177/00222429221102889.

[9] Z. Wang, S. Shen, Z. Wang, B. Chen, X. Chen and J. Wen, "Unbiased sequential recommendation with latent confounders," in *Proc. ACM Web Conf.*, Lyon, France, 2022, pp. 2195–2204. doi: 10.1145/3485447.3512092.

[10] Q. Li, X. Wang, Z. Wang, and G. Xu, "Be causal: De-biasing social network confounding in recommendation," *ACM Trans. Knowl. Discov. Data*, vol. 17, no. 1, pp. 1–23, Jan. 2023. doi: 10.1145/3533725.

[11] X. Yang, H. Zhang, G. Qi, and J. Cai, "Causal attention for vision-language tasks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Virtual Event, 2021, pp. 9847–9857. doi: 10.1109/CVPR46437.2021.00972.

[12] L. Chen, Y. Zheng, Y. Niu, H. Zhang, and J. Xiao, "Counterfactual samples synthesizing and training for robust visual question answering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13218–13234, Nov. 2023. doi: 10.1109/TPAMI.2023.3290012.

[13] Y. Liu, G. Li, and L. Lin, "Cross-modal causal relational reasoning for event-level visual question answering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 11624–11641, Oct. 2023. doi: 10.1109/TPAMI.2023.3284038.

[14] T. Wei, F. Feng, J. Chen, Z. Wu, J. Yi and X. He, "Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Min.*, Singapore, 2021, pp. 1791–1800. doi: 10.1145/3447548.3467289.

[15] D. Yu, Q. Li, X. Wang, and G. Xu, "Deconfounded recommendation via causal intervention," *Neurocomputing*, vol. 529, no. 10, pp. 128–139, Jun. 2023. doi: 10.1016/j.neucom.2023.01.089.

[16] Y. Zhu, J. Yi, J. Xie, and Z. Chen, "Deep causal reasoning for recommendations," *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 4, pp. 1–25, Oct. 2024. doi: 10.1145/3653985.

[17] S. Jamshidi *et al.*, "Effective text classification using BERT, MTM LSTM, and DT," *Data Knowl. Eng.*, vol. 151, no. 3, 2024, Art. no. 102306. doi: 10.1016/j.datak.2024.102306.

[18] M. Ghaderzadeh, A. Shalchian, G. Irajian, H. Sadeghsalehi, A. Z. Bialvaei and B. Sabet, "Artificial intelligence in drug discovery and development against antimicrobial resistance: A narrative review," *Iranian J. Med. Microbiol.*, vol. 18, no. 3, pp. 135–147, 2024. doi: 10.30699/ijmm.18.3.135.

[19] L. Wang, Z. He, R. Dang, M. Shen, C. Liu and Q. Chen, "Vision-and-language navigation via causal learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, 2024, pp. 13139–13150. doi: 10.48550/arXiv.2404.10241.
[20] J. Pearl, *Causality*. Cambridge, England: Cambridge University Press, 2009.
[21] Y. Jiang, R. Liang, J. Zhang, J. Sun, Y. Liu and Y. Qian, "Network public opinion detection during the coronavirus pandemic: A short-text relational topic model," *ACM Trans. Knowl. Discov. Data (TKDD)*, vol. 16, no. 3, pp. 1–27, May 2021. doi: 10.1145/3480246.
[22] B. Jiang, Z. Lu, N. Li, J. Wu, and Z. Jiang, "Retweet prediction using social-aware probabilistic matrix factorization," in *Proc. Int. Conf. On Comp. Sci.*, Wuxi, China, 2018, pp. 316–327. doi: 10.1007/978-3-319-93698-7_24.
[23] R. M. Safari, A. M. Rahmani, and S. H. Alizadeh, "Retweeting behavior prediction based on dynamic bayesian network classifier in microblogging networks," *Appl. Soft Comput.*, vol. 164, no. 2, 2024, Art. no. 111955. doi: 10.1016/j.asoc.2024.111955.
[24] L. Wang, Y. Zhang, J. Yuan, K. Hu, and S. Cao, "FEBDNN: Fusion embedding-based deep neural network for user retweeting behavior prediction on social networks," *Neural Comput. Appl.*, vol. 34, no. 16, pp. 13219–13235, 2022. doi: 10.1007/s00521-022-07174-9.
[25] D. Wang, W. Zhou, and S. Hu, "Information diffusion prediction with graph neural ordinary differential equation network," in *ACM Multimedia*, Melbourne, VIC, Australia: ACM SIGMM, 2024. doi: 10.1145/3664647.3681363.
[26] T. Wang, J. Huang, H. Zhang, and Q. Sun, "Visual commonsense R-CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 10760–10770. doi: 10.1109/CVPR42600.2020.01077.
[27] X. Zheng, B. Aragam, P. K. Ravikumar, and E. P. Xing, "Dags with no tears: Continuous optimization for structure learning," in *Adv. Neural Inf. Process. Syst.*, Montréal, QC, Canada, 2018, vol. 31.
[28] Y. He, P. Cui, Z. Shen, R. Xu, F. Liu and Y. Jiang, "DARING: Differentiable causal discovery with residual independence," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Min.*, Singapore, 2021, pp. 596–605. doi: 10.1145/3447548.3467439.
[29] Z. Wang, X. Gao, X. Liu, X. Rui, and Q. Zhang, "Incorporating structural constraints into continuous optimization for causal discovery," *Neurocomputing*, vol. 595, no. 3, 2024, Art. no. 127902. doi: 10.1016/j.neucom.2024.127902.
[30] Q. Xie, X. Zhang, Y. Ding, and M. Song, "Monolingual and multilingual topic analysis using LDA and BERT embeddings," *J. Informetr.*, vol. 14, no. 3, 2020, Art. no. 101055. doi: 10.1016/j.joi.2020.101055.
[31] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, CA, USA, 2016, pp. 855–864. doi: 10.1145/2939672.2939754.
[32] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *Int. Conf. Mach. Learn.*, Lille, France, 2015, pp. 2048–2057.
[33] A. Vaswani, "Attention is all you need," 2017, *arXiv:1706.03762*.
[34] Q. Zhang, Y. Gong, J. Wu, H. Huang, and X. Huang, "Retweet prediction with attention-based deep neural network," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Indianapolis, IN, USA, 2016, pp. 75–84. doi: 10.1145/2983323.298380.
[35] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, vol. 30.
[36] S. Deng, H. Rangwala, and Y. Ning, "Learning dynamic context graphs for predicting social events," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, Anchorage, AK, USA, 2019, pp. 1007–1016. doi: 10.1145/3292500.3330919.
[37] Z. Zhao *et al.*, "Attentional image retweet modeling via multi-faceted ranking network learning," in *Proc. Twenty-Seventh Int. Joint Conf. Artifici. Intellig.*, Stockholm, Sweden, 2018, pp. 3184–3190. doi: 10.24963/ijcai.2018/442.
[38] H. Yin, S. Yang, X. Song, W. Liu, and J. Li, "Deep fusion of multimodal features for social media retweet time prediction," *World Wide Web*, vol. 24, no. 4, pp. 1027–1044, 2021. doi: 10.1007/s11280-020-00850-7.

[39] W. J. Sun, X. F. Liu, and F. Shen, "Learning dynamic user interactions for online forum commenting prediction," in *Proc. IEEE Int. Conf. Data Min.*, Auckland, New Zealand, 2021, pp. 1342–1347. doi: 10.1109/ICDM51629.2021.00168.

[40] X. Ma and Y. Huo, "Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework," *Technol. Soc.*, vol. 75, no. 5, 2023, Art. no. 102362. doi: 10.1016/j.techsoc.2023.102362.

[41] S. Bhagat and D. J. Kim, "Examining users' news sharing behaviour on social media: Role of perception of online civic engagement and dual social influences," *Behav. Inf. Technol.*, vol. 42, no. 8, pp. 1194–1215, 2023. doi: 10.1080/0144929X.2022.2066019.

[42] B. Li, S. W. Dittmore, O. K. M. Scott, W. -J. Lo, and S. Stokowski, "Why we follow: Examining motivational differences in following sport organizations on Twitter and Weibo," *Sport Manag. Rev.*, vol. 22, no. 3, pp. 335–347, 2019. doi: 10.1016/j.smr.2018.04.006.

[43] A. Hapfelmeier, R. Hornung, and B. Haller, "Efficient permutation testing of variable importance measures by the example of random forests," *Comput. Stat. Data Anal.*, vol. 181, 2023, Art. no. 107689. doi: 10.1016/j.csda.2022.107689.

[44] I. Khemakhem, R. Monti, R. Leech, and A. Hyvarinen, "Causal autoregressive flows," in *Proc. Int. Conf. Artif. Intell. Stat.*, vol. 130, pp. 3520–3528, 2021.