mathematics

Article

Temporal Graph Attention Network for Spatio-Temporal Feature Extraction in Research Topic Trend Prediction

Zhan Guo [1], Mingxin Lu [2,3,*] and Jin Han [1]

[1] School of Software, Nanjing University of Information Science and Technology, Nanjing 210044, China; 202212490409@nuist.edu.cn (Z.G.); 002254@nuist.edu.cn (J.H.)
[2] Department of Information Management, Nanjing University, Nanjing 210044, China
[3] Nanjing University (Suzhou) High-Tech Institute, Suzhou 215000, China
* Correspondence: mxlu@nju.edu.cn

Abstract: Comprehensively extracting spatio-temporal features is essential to research topic trend prediction. This necessity arises from the fact that research topics exhibit both temporal trend features and spatial correlation features. This study proposes a Temporal Graph Attention Network (T-GAT) to extract the spatio-temporal features of research topics and predict their trends. In this model, a temporal convolutional layer is employed to extract temporal trend features from multivariate topic time series. Additionally, a multi-head graph attention layer is introduced to capture spatial correlation features among research topics. This layer learns attention scores from the data by using scaled dot product operations and updates edge weights between topics accordingly, thereby mitigating the issue of over-smoothing. Furthermore, we introduce WFtopic-econ and WFtopic-polit, two domain-specific datasets for Chinese research topics constructed from the Wanfang Academic Database. Extensive experiments demonstrate that T-GAT outperforms baseline models in prediction accuracy, with RMSE and MAE being reduced by 4.8% to 7.1% and 14.5% to 18.4%, respectively, while R^2 improved by 4.8% to 7.9% across varying observation time steps on the WFtopic-econ dataset. Moreover, on the WFtopic-polit dataset, RMSE and MAE were reduced by 4.0% to 5.3% and 10.0% to 10.7%, respectively, and R^2 improved by 7.6% to 14.4%. These results validate the effectiveness of integrating graph attention with temporal convolution to model the spatio-temporal evolution of research topics, providing a robust tool for scholarly trend analysis and decision making.

Keywords: topic trend prediction; feature extraction; graph attention; temporal convolution; multivariate time-series forecasting

MSC: 68T01

check for updates

Academic Editor: Jüri Majak

Received: 26 December 2024
Revised: 11 February 2025
Accepted: 17 February 2025
Published: 20 February 2025

Citation: Guo, Z.; Lu, M.; Han, J. Temporal Graph Attention Network for Spatio-Temporal Feature Extraction in Research Topic Trend Prediction. *Mathematics* **2025**, *13*, 686. https://doi.org/10.3390/math13050686

1. Introduction

Research topic trend prediction is an important task in time-series forecasting. Predicting trends in research topics from a vast number of academic studies can significantly enhance scholars' understanding of prospective research directions and facilitate the advance planning of research projects. Furthermore, it can assist journals in formulating publishing strategies and improving the quality of their publications [1,2]. Accurately extracting the spatio-temporal features of research topics is essential to predicting their trends. This is due to the fact that research topics exhibit both temporal trend features and spatial correlation features. Specifically, the popularity of research topics evolves over time, and the popularity of a particular research topic is often associated with the popularity of related topics [3].

Traditional methods typically rely on recurrent neural networks (RNNs) to extract temporal trend features. These approaches often treat each research topic as an independent entity, assigning a separate RNN to each topic. However, this strategy overlooks the extraction of spatial correlation features, and the number of RNNs increases with the number of research topics. Lately, methods based on spatio-temporal graph neural networks incorporate the extraction of spatial correlation features through ensemble graph convolutional networks (GCNs). Nevertheless, GCNs [4] assume that the influence between neighboring nodes is fixed, which can result in over-smoothing [5], where the features of nodes converge to similar values after multiple rounds of neighborhood aggregation. To address these limitations, this paper proposes a Temporal Graph Attention Network (T-GAT) designed to comprehensively extract the spatio-temporal features of research topics.

In the proposed framework, we introduce temporal convolution to simultaneously extract temporal trend features from multiple research topics. Temporal convolution can process multivariate time series in parallel, and by doubling the dilation coefficient layer by layer, the receptive field of the network can grow exponentially, enabling it to manage longer historical information. Additionally, we propose a multi-head graph attention layer based on scaled dot product to extract spatial correlation features among research topics. Multi-head graph attention can learn attention scores from the data and adjust the weights of neighboring nodes based on these scores during neighborhood aggregation, thereby mitigating over-smoothing. We then integrate these two components to achieve the fusion of temporal and correlation features.

In summary, the proposed method addresses the following challenges in research topic trend prediction tasks: (1) the limitation of low prediction accuracy due to the lack of the extraction of correlation features among research topics and (2) the tendency for over-smoothing that occurs when GCNs extract correlated features. The main contributions of this study are as follows:

1. We propose a multi-head graph attention layer based on scaled dot product to extract correlation features among research topics. This layer can learn attention scores from the data and adjust the weights of neighboring nodes accordingly during neighborhood aggregation;
2. We integrate graph attention with temporal convolution to propose a Temporal Graph Attention Network (T-GAT) that effectively extracts both temporal trend features and spatial correlation features of research topics. The fusion of spatio-temporal features enhances predictive performance;
3. We constructed research topic datasets for economics and politics, referred to as WFtopic-econ and WFtopic-polit, by using the metadata from papers in the Wanfang Chinese Academic Database. After conducting extensive experiments and analyses on the datasets, the method proposed in this paper demonstrated superior performance compared with the baseline method.

2. Related Works

Research topic trend prediction is a specialized domain within time-series forecasting that requires modeling both temporal dynamics and spatial dependencies among interconnected topics. This section reviews recent advances in traditional time-series forecasting methods and spatio-temporal graph neural networks (STGNNs), focusing on their strengths, limitations, and relevance to research topic trend analysis.

2.1. Traditional Time-Series Forecasting Methods

Traditional approaches to time-series forecasting fall into two categories: statistical models and neural network-based methods.

Classical statistical techniques, such as Autoregressive Integrated Moving Average (ARIMA) and Prophet, have long been used for trend prediction due to their simplicity and interpretability. ARIMA models capture linear relationships between past and future values through differencing and autoregressive components [6]. Prophet, developed by Facebook, extends this by incorporating seasonality and holiday effects, making it suitable for datasets with periodic patterns [7]. For instance, Zou et al. [8] applied ARIMA to predict the trends of the five most popular research topics in Chinese policy research papers, achieving moderate accuracy. Yu et al. [9] detected emerging scientific topics by using Neural Prophet to predict emerging attributes. However, these models struggle with non-linear patterns and multivariate data, which are inherent in research topic trends where multiple topics interact dynamically.

The rise of deep learning shifted the focus toward recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), which excel at capturing temporal dependencies. For example, Yang et al. [10] used LSTM to predict the emerging index of research topics in neoplasms and metabolism domains by treating each topic as an independent time series. Zhang et al. [11] introduced EEMD to decompose complex time series into simple subsequences on key technical topics in aircraft assembly and used a GRU to predict the trends of each subsequence separately. Ma et al. [12] introduced a time-masked selection mechanism to minimize redundant information in time-series data. While RNNs have improved accuracy over ARIMA, they suffer from scalability issues: training separate RNNs for thousands of topics is computationally prohibitive. Additionally, RNNs inherently process sequences sequentially, leading to slow training times and difficulty in parallelization [13].

Temporal Convolutional Networks (TCNs) [14] emerged as a competitive alternative by leveraging dilated convolutions to model long-term dependencies efficiently. Unlike RNNs, TCNs process entire sequences in parallel, enabling faster training and larger receptive fields. Gopali et al. [15] demonstrated that TCNs outperform LSTM networks in multivariate time-series tasks due to their ability to expand receptive fields exponentially through stacked dilated layers. Li et al. [16], Ye et al. [17], and Xue et al. [18] applied their improved TCNs to predict ship traffic flow, ride-hailing demand, and nitrogen replacement gas volume changes, respectively. These works show robustness in handling noisy, high-dimensional data. Despite these advances, standalone TCNs ignore spatial correlations among variables (e.g., related research topics), limiting their utility in interconnected systems.

2.2. Spatio-Temporal Graph Neural Networks

To address the limitations of isolated temporal modeling, spatio-temporal graph neural networks (STGNNs) integrate graph structures to capture spatial correlations and temporal dynamics simultaneously.

GCNs encode spatial dependencies by aggregating features from neighboring nodes in a graph. Early applications in traffic prediction, such as Zhao et al.'s [19] T-GCN, combined GCNs for road network topology and GRUs for temporal trends. In academic trend prediction, Geng et al. [20] constructed a dynamic heterogeneous graph of research topics, papers, authors, and venues, using GCNs to propagate influence among connected nodes. While these methods improved accuracy by incorporating spatial relationships, traditional GCNs assume fixed edge weights during neighborhood aggregation, leading to two critical issues: first, over-smoothing, as repeated graph convolutions cause node features to converge to similar values, erasing discriminative patterns [5]; second, static relationships, as GCNs are unable to adapt to evolving spatial dependencies, such as shifting topic influences over time [21].

To mitigate over-smoothing and enable dynamic spatial modeling, graph attention networks (GATs) were proposed. GATs assign learnable attention scores to neighbors, allowing the model to focus on relevant nodes during aggregation. Veličković et al. [22] introduced multi-head attention in GATs, enabling richer representations by aggregating information from multiple attention heads. This approach has been widely adopted in traffic prediction (e.g., Fan et al.'s [23] RGDAN) and social network analysis (e.g., Wang et al.'s [24] RLGAT). For academic trends, Zou et al. [25] applied GATs to model the mutual influence of papers within citation networks, successfully achieving citation link prediction. However, most GAT-based methods focus on static graphs and do not integrate temporal modeling, leaving a gap for joint spatio-temporal frameworks.

Recent works combine GNNs with temporal modules to jointly learn spatial and temporal features. For instance, Graph WaveNet [26] integrates dilated TCNs with diffusion GCNs to capture long-term temporal dependencies and dynamic spatial relationships in traffic prediction. Similarly, Peng et al. [27] proposed a GTRGAT for intrusion detection in the Industrial Internet of Things (IIoT). This model utilizes GATs to model spatial characteristics of devices and gated TCNs for temporal features. These hybrid models have achieved state-of-the-art results in physical systems, such as traffic and the Internet of Things; however, they remain underexplored in academic trend prediction.

3. Problem Definition

In this study, the research topics are represented by the keywords from papers. The research topics are conceptualized as nodes, while the co-occurrence relationships among them are represented as edges. Consequently, the research topics can be modeled as a graph $G = (V,\ E)$, where V is the set of nodes, with the number of nodes being $|V| = n$, and E is the set of edges. The structural representation of the graph is maintained in a weighted adjacency matrix $A \in\ \mathbb{R}^{n\times n}$, where element $A_{i,j}$ indicates the co-occurrence frequency of research topics v_i and v_j across all keyword lists. Then, the popularity of a research topic within a specific year is quantified by the aggregate citation count of all papers pertaining to that topic during that year. Let $x_i^t \in \mathbb{R}^d$ denote the features of node i at time t; then, matrix $X^t = \{x_1^t, x_2^t, \ldots,\ x_n^t\} \in \mathbb{R}^{n\times d}$ encapsulates the features of all nodes at time t. Therefore, the objective of the research topic trend prediction problem is to learn a function f capable of forecasting the popularity of each research topic for the subsequent l time steps, utilizing the provided research topic graph G and the popularity data from the preceding p time steps. The mapping function follows the formulation in [26]:

$$\hat{X}^{(t+1):(t+l)} = f\left(X^{(t-p):t}, G\right), \tag{1}$$

where $X^{(t-p):t} \in \mathbb{R}^{p\times n\times d}$ and $\hat{X}^{(t+1):(t+l)} \in \mathbb{R}^{l\times n\times d}$.

4. Method

This section first introduces the details of the temporal feature extraction module and the correlation feature extraction module and then introduces the overall architecture of the proposed model.

4.1. Temporal Trend Feature Extraction

In the temporal dimension, fluctuations in the popularity of individual research topics reveal distinct trend features. Inspired by Graph WaveNet [26], we employ Temporal Convolutional Networks (TCNs) to extract temporal trend features of the research topics. In this module, we remove the gating mechanism to reduce parameter complexity and add weight normalization for the weight parameters. This approach mitigates gradient fluctuations during training, thereby facilitating a smoother optimization process and enhancing the model's convergence. Furthermore, residual connections are incorporated following each convolution operation to address the issue of gradient vanishing in deep networks.

TCNs offer significant advantages over recurrent neural networks (RNNs). Firstly, TCNs do not exhibit temporal step dependency when processing time-series data, thereby facilitating the use of parallel computing to enhance computational efficiency. Secondly, the dilated causal convolution within TCNs significantly improves the model's ability to capture long-term dependencies.

Figure 1 illustrates the fundamental principle of temporal convolution. The essence of the TCN is rooted in dilated causal convolution. Causal convolution, in particular, guarantees that the convolution operation is influenced solely by the current and preceding time steps and is not affected by information from future time steps. This characteristic is essential to adhering to the causal requirements inherent in time-series data. Dilated causal convolution introduces intervals between the elements of the kernel, thereby enabling the kernel to execute causal convolution over extended time intervals. Specifically, if we denote the feature representation of a research topic at time t by x_t, the output of the TCN at time t can be mathematically expressed following the formulation in [14]:

$$x'_t = \sum_{i=0}^{k-1} W_i x_{t-d\cdot i} \tag{2}$$

where W_i is the weight of the i-th convolution kernel, k is the kernel size, and d is the dilation coefficient. By incrementally doubling the dilation coefficient at each layer, the TCN can attain a larger receptive field with fewer layers, thereby effectively capturing long-term temporal dependencies.

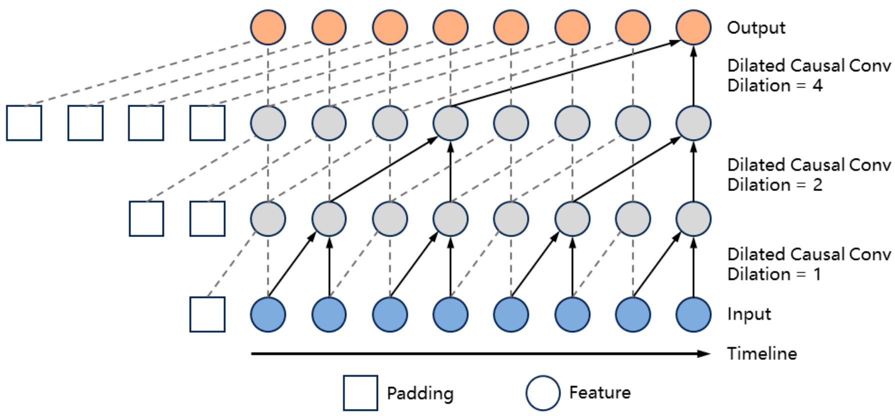


Figure 1. Temporal convolution with a kernel size of 2.

4.2. Spatial Correlation Feature Extraction

In the spatial dimension, the popularity of a given research topic is often associated with the popularity of related topics. The precise extraction of this spatial correlation feature is crucial to enhancing prediction accuracy. In this study, we propose a multi-head graph attention (GAT) layer based on scaled dot product to extract spatial correlation features among research topics. Unlike the static edge weights employed in Graph Convolutional Networks (GCNs), the edge weights in GAT during neighborhood aggregation are determined by dynamically learned attention scores derived from the data, thereby mitigating the issue of over-smoothing.

Figure 2 illustrates the multi-head graph attention mechanism proposed in this study, which is implemented in a manner distinct from the widely applied Veličković et al.'s GAT [22]. They utilize additive attention, which computes attention scores through linear transformations and pairwise summation. In contrast, we utilize scaled dot product attention, a method derived from the Transformer architecture [28]. This approach calculates attention scores through dot product and scaling operations, which can be mathematically expressed following the formulation in [28]:

$$S = \frac{XW_q(XW_k)^T}{\sqrt{d_k}} \tag{3}$$

where $X \in \mathbb{R}^{n \times d}$ is the feature representation of all nodes and matrices $W_q \in \mathbb{R}^{d \times d_k}$ and $W_k \in \mathbb{R}^{d \times d_k}$ represent the linear transformations for the query and key, respectively. The term $\sqrt{d_k}$ serves as a scaling factor, where d_k denotes the hidden dimension. This scaling factor is employed to mitigate the numerical instability issues that may arise due to the increase in dimensionality.

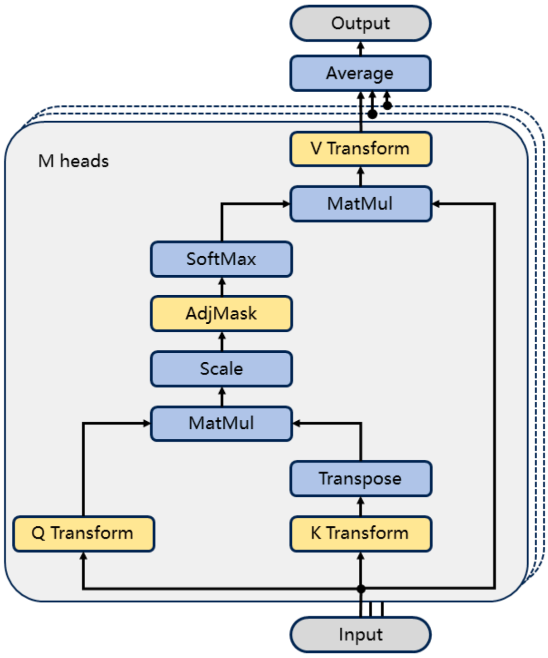


Figure 2. Multi-head graph attention mechanism based on scaled dot product.

In our work, we have taken into account the structural information provided by the graph and proposed a masking operation on the attention scores S based on adjacency matrix A. This can be expressed by using the following equation:

$$S_{ij} = \begin{cases} S_{ij}, & A_{ij} \neq 0 \\ -9 \times 10^{15}, & A_{ij} = 0 \end{cases} \tag{4}$$

where -9×10^{15} is a real number that approaches negative infinity. The objective of this approach is to ensure that the attention scores of non-neighboring nodes are effectively reduced to zero following softmax normalization. This masking operation restricts the model's focus exclusively to neighboring nodes.

After applying the $softmax$ function to normalize the attention scores, we aggregate the features of neighboring nodes in the graph according to these scores to obtain the updated feature representations for all nodes. This process can be mathematically expressed as follows:

$$X' = softmax(S)XW_v + b \quad (5)$$

where $W_v \in \mathbb{R}^{d \times d_v}$ and $b \in \mathbb{R}^{n \times d_v}$ represent the linear transformation parameters and biases associated with the value vectors, respectively.

In the multi-head graph attention mechanism, M distinct attention heads independently compute attention scores and aggregate the features of neighboring nodes. Subsequently, the outputs of all attention heads are averaged to derive the final feature representation for all nodes. This process can be mathematically expressed as follows:

$$X' = \frac{1}{M} \sum_{m=1}^{M} (softmax(S^m)XW_v^m + b^m) \quad (6)$$

The multi-head attention graph mechanism not only captures the intricate correlations among research topics but also enhances the model's robustness and generalization capabilities.

4.3. Model Architecture

Figure 3 illustrates the overall architecture of the Temporal Graph Attention Network (T-GAT) proposed in this study, which comprises multiple stacked spatio-temporal feature extraction blocks and an output section. Each spatio-temporal block includes a temporal convolutional layer and a multi-head graph attention layer, which are utilized to extract temporal trend features and spatial correlation features pertinent to various research topics, respectively. The output from each layer undergoes a residual connection to mitigate the issue of gradient vanishing, as well as layer normalization to adjust the data distribution, followed by activation through the ReLU activation function. A 1×1 convolution is employed to align the hidden dimensions of the residuals with the output of the current layer.

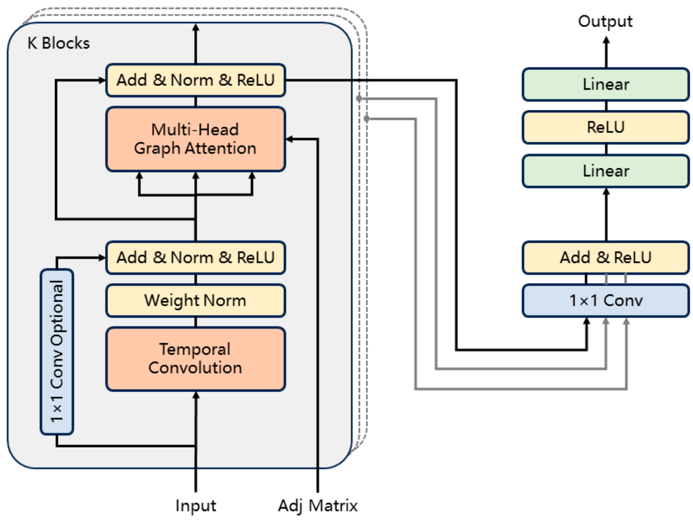


Figure 3. Architecture of Temporal Graph Attention Network.

It is noteworthy that the weight normalization applied by the temporal convolutional layer pertains specifically to the network's weight parameters rather than the data themselves, in order to mitigate gradient fluctuations during training. The dilation coefficient of the temporal convolutional layer is doubled with each passage through a spatio-temporal block from the bottom to the top, enabling the lower layers to address short-term trends while the upper layers manage long-term trends. Concurrently, the graph attention layer aggregates the features of various neighboring nodes based on attention scores. Ultimately, the output section is composed of two fully connected layers.

5. Experiments and Discussion

5.1. Dataset

In this study, we constructed research topic datasets for economics and politics, referred to as WFtopic-econ and WFtopic-polit, by using the metadata from papers in the Wanfang Chinese Academic Database.

For WFtopic-econ, we first collected metadata from 40,254 papers in the field of economics, spanning the years 1999 to 2021. Each metadata entry includes various elements, such as title, abstract, keywords, publication year, and citation count. We then extracted all keywords that appeared at least 20 times in the metadata collection to serve as the final research topics, which represent the nodes of the graph. Subsequently, we constructed an adjacency matrix based on the co-occurrence relationships between research topics, which represent the edge weights of the graph. The value of each element in the adjacency matrix indicates the co-occurrence frequency of a corresponding pair of research topics across all keyword lists in the relevant literature. Concurrently, we constructed a multivariate time series based on the annual citation count of each research topic, representing the dynamic attributes of the nodes in the graph. In this matrix, each column corresponds to a specific research topic, each row represents a specific year, and the values in the matrix indicate the total number of citations for all papers associated with that topic in the corresponding year. The construction process for WFtopic-polit follows a similar methodology.

Basic information regarding the research topic datasets from these two distinct disciplines is presented in Table 1, which provides a foundation for the subsequent experimental analysis.

Table 1. Basic information of datasets.

Dataset	Nodes	Edges	Time Span	Min	Max	Standard Deviation
WFtopic-econ	367	6212	1991–2021	0	4203	234.697
WFtopic-polit	146	1932	1991–2021	0	3010	168.088

5.2. Experimental Setup

In this article, the samples and labels of the dataset are generated by segmenting the original time-series data by using a sliding window approach. Specifically, an observation window with a time step of 4 and a prediction window with a time step of 1 are employed to traverse the time sequence. The values contained within the observation window serve as samples, while the values within the prediction window are designated as labels. Subsequently, the dataset is divided in chronological order, with 60% being allocated to the training set and 40% to the test set.

During the model training process, it is recommended to set the number of training epochs to 3000 and to implement an early stopping strategy. Specifically, training should be halted if the loss does not exhibit a decrease over 50 consecutive training iterations, thereby mitigating the risk of overfitting. The loss function employed is the Mean Squared Error (MSE), while the Adam optimizer is utilized with a learning rate of 1×10^{-3} and a weight decay of 1×10^{-5}. The model's hidden dimension is set to 32, and the convolution kernel size for the temporal convolutional layer is set to 2. The dilation coefficient is designed to double with each increase in the number of spatio-temporal blocks, commencing from a value of 1. To ensure that the input and output time lengths are equal, padding and truncation operations are applied. Furthermore, the number of attention heads in the multi-head graph attention layer is established as 4.

5.3. Evaluation Metrics

To conduct a thorough assessment of the accuracy of the model's predictions regarding research topic trends, this paper employs several evaluation metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R-squared, R^2). These metrics provide a multifaceted evaluation of the model's predictive performance, thereby enhancing the comprehensiveness and reliability of the assessment outcomes.

RMSE serves as an indicator of the overall prediction error, with smaller values indicating greater predictive accuracy of the model. The formula is expressed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

where y_i denotes the actual value of the i-th sample, while $\hat{y}_i$ signifies the predicted value of the i-th sample. Additionally, n represents the number of samples.

MAE quantitatively represents the average discrepancy between predicted values and actual values; thus, a smaller MAE indicates a higher level of predictive accuracy of the model. The formula is expressed as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

R^2 serves as a metric for assessing the degree of fit between a model and the corresponding data and is expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{9}$$

where $\overline{y}$ is the mean of the actual values. The range of R^2 is from 0 to 1: a value closer to 1 indicates a stronger explanatory power of the model with respect to the data.

5.4. Results and Discussion

5.4.1. Performance Evaluation

In order to conduct a comprehensive evaluation of the performance of the T-GAT model proposed in this article, this section compares it with several widely utilized methods for time-series forecasting and spatio-temporal graph neural networks:

- LSTM: A variant of RNNs designed to capture temporal dependencies through the incorporation of memory cells and gating mechanisms;
- GRU: A variant of RNNs that regulates the transmission and updating of information via gating mechanisms. The structure of GRU is less complex compared with that of LSTM;
- T-GCN [19]: It integrates GCNs and GRUs to effectively capture both spatial and temporal features of the data simultaneously;
- Graph WaveNet (GWNet) [26]: It integrates GCNs and TCNs to effectively capture both spatial and temporal features of the data.

Table 2 presents a comparative analysis of the performance of T-GAT and various baseline models in predicting research topic trends in two datasets. In the WFtopic-econ dataset, the results indicate that the RMSE of spatio-temporal graph neural networks, including T-GAT, T-GCN, and Graph WaveNet, is between 10.75 and 43.16 lower than that of traditional time-series forecasting models such as LSTM and GRU. This finding underscores the importance of extracting spatial correlation features of research topics for enhancing prediction performance. Moreover, the MAE of Graph WaveNet, which employs TCNs for the extraction of temporal trend features, is found to be between 6.14 and 14.60 lower than that of T-GCN, which utilizes GRUs for temporal feature extraction. This observation highlights the superior feature extraction capabilities of convolutional architectures in comparison to recurrent architectures. Additionally, the MAE of the T-GAT proposed in this study is between 10.60 and 14.63 lower than that of Graph WaveNet, further demonstrating the advantages of T-GAT in effectively extracting the spatio-temporal features of research topics. Similarly, in the WFtopic-polit dataset, the RMSE and MAE of T-GAT are the lowest, ranging from 86.209 to 97.213 and from 51.820 to 58.256, respectively, while the Coefficient of Determination (R^2) is the highest, ranging from 0.438 to 0.465.

Figure 4 illustrates the variation in MAE and R^2 for each method across different observation time steps. It is evident that as the observation time step increases, the MAE for each model exhibits a downward trend, suggesting that an extended observation history enhances the model's learning capabilities. Specifically, in the WFtopic-econ dataset, the T-GAT proposed in this study requires only three to five observation time steps to achieve a significant reduction in the MAE of the predictive outcomes. This finding implies that to forecast the popularity of research topics in the field of economics, it is advisable to utilize historical data from the past 3 to 5 years to obtain satisfactory results. Conversely, in the WFtopic-polit dataset, longer observation time steps correlate with lower MAE. This indicates that in the political field, it is recommended to use the longest possible historical data to predict trends in research topics. In both datasets, the R^2 value of T-GAT remains more stable across varying time steps compared with other models, indicating that T-GAT demonstrates superior robustness in data interpretation. Furthermore, the results of the *t*-tests for T-GAT and Graph WaveNet are illustrated in Figure 4, with all *p*-values being less than 0.05, confirming the improvements of T-GAT.

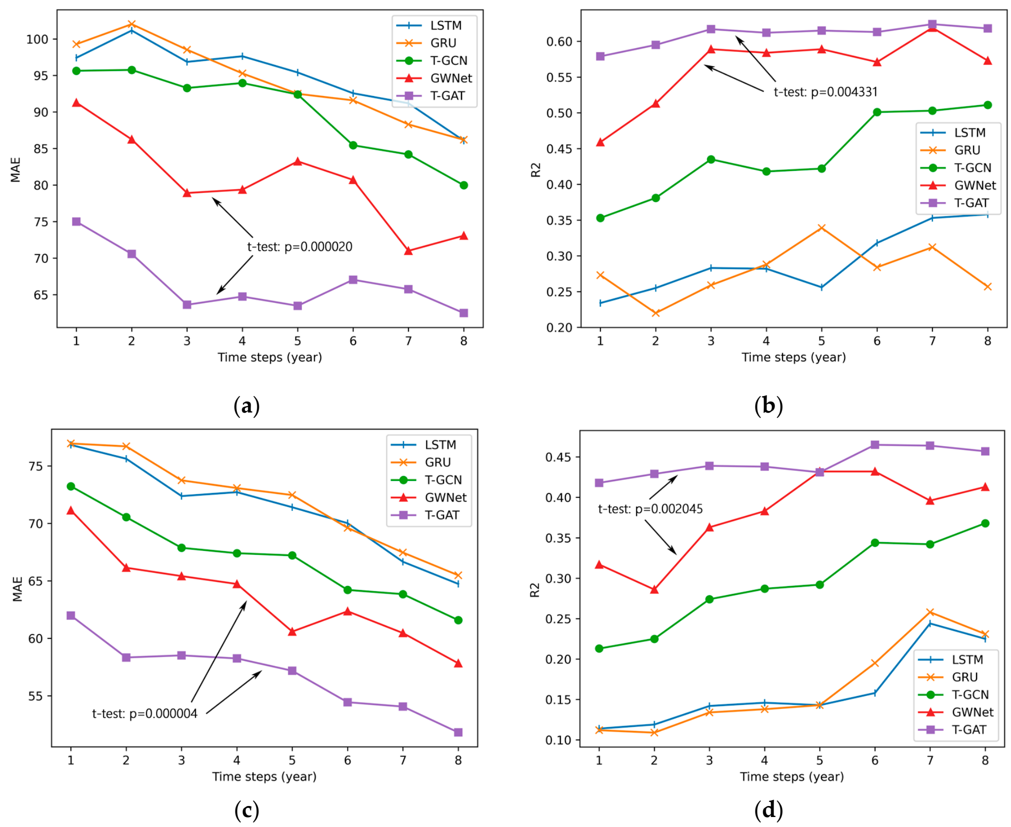


Figure 4. Performance variation across different observation time steps: (**a**) Mean Absolute Error on WFtopic-econ; (**b**) R-squared on WFtopic-econ; (**c**) Mean Absolute Error on WFtopic-polit; (**d**) R-squared on WFtopic-polit.

Table 2. Performance comparison of T-GAT and other baseline models.

Dataset	Model	Time Steps = 4			Time Steps = 6			Time Steps = 8		
		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
WFtopic-econ	LSTM	156.79	97.63	0.282	149.60	92.57	0.318	138.98	86.07	0.358
	GRU	155.25	95.28	0.288	152.37	91.59	0.284	147.81	86.19	0.257
	T-GCN	144.15	93.97	0.418	136.60	85.45	0.501	128.65	79.97	0.511
	GWNet	119.33	79.37	0.584	120.05	80.71	0.571	112.94	73.07	0.573
	T-GAT	113.63	64.74	0.612	111.54	67.04	0.613	105.32	62.47	0.618
WFtopic-polit	LSTM	119.790	72.721	0.146	114.368	70.022	0.158	105.079	64.741	0.225
	GRU	120.750	73.076	0.138	115.073	69.612	0.195	106.879	65.489	0.231
	T-GCN	108.227	67.412	0.287	103.276	64.216	0.344	100.159	61.587	0.368
	GWNet	102.609	64.733	0.383	94.897	62.360	0.432	89.963	57.834	0.413
	T-GAT	97.213	58.256	0.438	91.116	54.446	0.465	86.209	51.820	0.457

Table 3 shows the training efficiency of T-GAT and other baseline models. In both datasets, GRU exhibits the shortest training time, while the LSTM model has a slightly longer training duration than GRU. The training times for spatio-temporal graph neural networks, such as T-GAT, T-GCN, and Graph WaveNet, are 0.021 to 0.036 s per epoch longer than that of GRU. This suggests that the integration of GNNs to extract the correlation features of research topics introduces additional computational complexity, resulting in increased time overhead. Among these models, T-GAT has the longest training time, at 0.054 s per epoch and 0.035 s per epoch. This may be attributed to the high computational complexity associated with multi-head graph attention. This limitation will be further analyzed and addressed in future work. Additionally, T-GAT demonstrates the fastest convergence speed, achieving convergence in 31 epochs and 45 epochs. This indicates that the correlation features extracted through multi-head graph attention can accelerate the model's convergence.

Table 3. Training efficiency comparison of T-GAT and other baseline models.

Dataset	Model	Training Time (s/epoch)	Convergence Speed (epochs)
WFtopic-econ	LSTM	0.018	1032
	GRU	0.017	926
	T-GCN	0.039	633
	GWNet	0.050	48
	T-GAT	0.054	31
WFtopic-polit	LSTM	0.008	1068
	GRU	0.006	939
	T-GCN	0.027	774
	GWNet	0.031	51
	T-GAT	0.035	45

5.4.2. Interpretability Analysis

The experiments and visualization presented in this section are based on the WFtopic-econ dataset, which encompasses Chinese research topics within the field of economics. The topics involved were translated into English in the visualization.

In order to evaluate the effectiveness of the model proposed in this paper more intuitively, the actual values and the prediction curves of T-GAT and Graph WaveNet (GWNet) for four randomly selected research topics are illustrated in Figure 5. The observation time step is set to 4 years, while the prediction time step is 1 year. The results indicate that the prediction curve of Graph WaveNet is relatively smooth, whereas T-GAT more accurately captures the trends in research topic changes. This discrepancy may be attributed to the over-smoothing phenomenon associated with the GCN utilized in Graph WaveNet, which tends to homogenize node features as the number of network layers increases. In contrast, T-GAT's ability to capture node correlations is grounded in an attention mechanism that dynamically adjusts the weights of different neighboring nodes during each neighborhood aggregation, thereby mitigating the risk of over-smoothing.

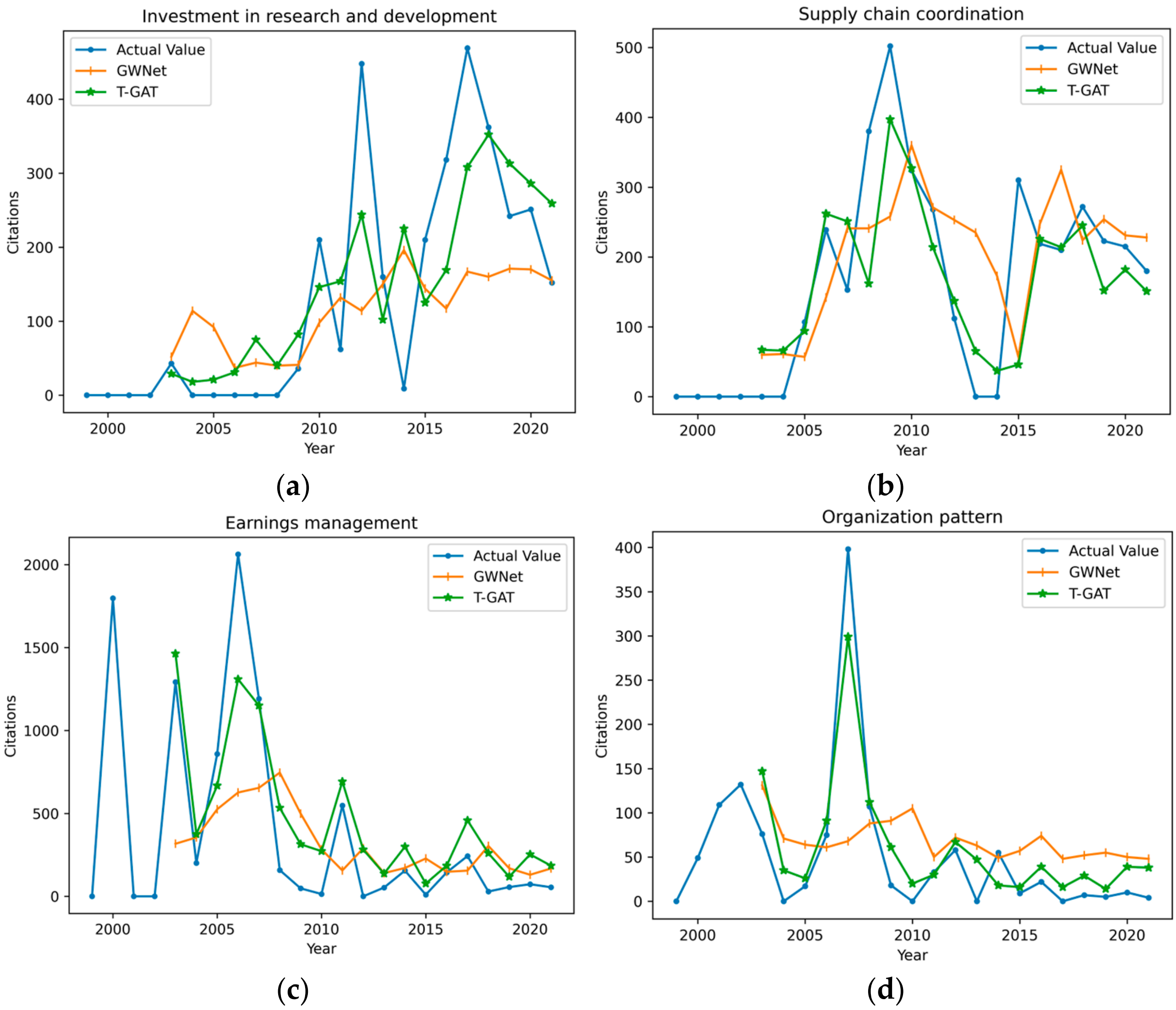
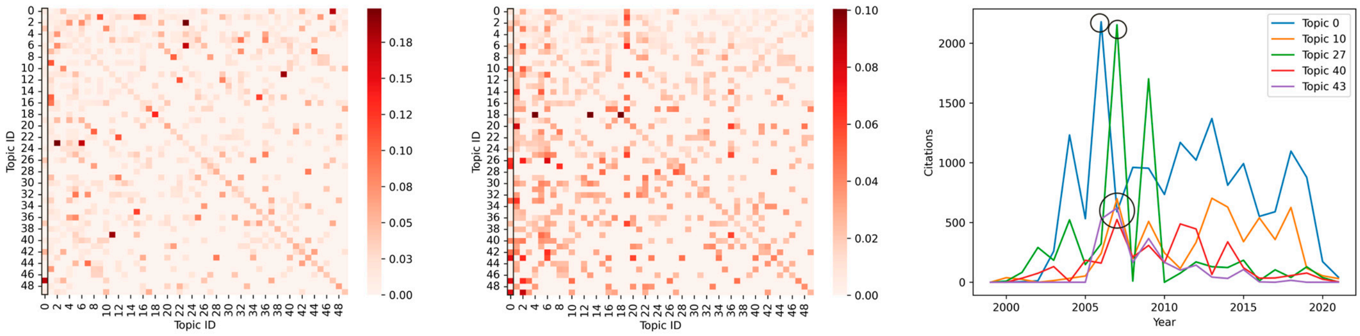


Figure 5. Comparison of prediction curves between T-GAT and Graph WaveNet for four research topics: (**a**) investment in research and development; (**b**) supply chain coordination; (**c**) earnings management; (**d**) organization pattern.

Figure 6 illustrates the capacity of the graph attention mechanism to extract correlation features of research topics. Specifically, panel (a) presents the standardized adjacency matrix of the top 50 research topics, panel (b) displays the attention matrix of these topics as learned by the model, and panel (c) provides a comparative analysis of the actual trends across multiple topics. It is noteworthy that the zeroth column of the attention matrix indicates that topic 0 possesses a higher weight relative to topics 10, 27, 40, 43, and others. This observation is corroborated in panel (c), where the citation count for topic 0 reached its zenith in 2006, while the corresponding topics 10, 27, 40, and 43 also experienced peak citation counts in the subsequent year. This suggests that the prominence of topic 0 exerts a significant influence on the aforementioned topics. Conversely, the adjacency matrix fails to capture this correlation, thereby highlighting the importance of dynamically adjusting the weights of neighboring nodes through the attention mechanism to more accurately extract the correlation features among research topics.



(**a**) (**b**) (**c**)

Figure 6. The correlation features identified by the graph attention mechanism: (**a**) Heatmap representing the adjacency matrix of the top 50 research topics. (**b**) Heatmap representing the attention matrix of the top 50 research topics. (**c**) Actual trends observed across multiple topics.

6. Conclusions

This study models multiple research topics as a graph and introduces a Temporal Graph Attention Network to predict their trends. In this framework, we utilize temporal convolution to extract the temporal features of each research topic and propose a multi-head graph attention layer based on scaled dot product to extract the correlation features among research topics. We then integrate these two components to achieve the fusion of temporal and correlation features.

We constructed research topic datasets for economics and politics, referred to as WFtopic-econ and WFtopic-polit, by using metadata from papers in the Wanfang Chinese Academic Database. Experiments conducted on both datasets indicate that the proposed model's extraction of correlation features effectively reduces prediction errors. Furthermore, the integration of multi-head graph attention demonstrates greater accuracy in predicting peak values compared with the graph convolution-based method.

However, the proposed model has certain limitations. On one hand, the training time of the model is relatively long. On the other hand, the model's ability to extract correlation features requires improvement.

In future research, we will concentrate on further enhancing the proposed model's ability to capture correlation features while also reducing its training time. Additionally, we will investigate how quantization methods, such as logarithmic quantization [29], affect model performance from the perspective of optimization techniques.

Author Contributions: Conceptualization, Z.G., M.L. and J.H.; methodology, Z.G.; software, Z.G.; validation, Z.G., M.L. and J.H.; formal analysis, M.L. and J.H.; investigation, Z.G.; resources, J.H.; data curation, Z.G.; writing—original draft preparation, Z.G.; writing—review and editing, M.L. and J.H.; visualization, Z.G.; supervision, M.L. and J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by "Design, Evaluation, and Research & Development of Lightweight Cryptographic Systems for the Application Scenarios of Blockchain" (SYC2022093), the key research and development project of Suzhou Science and Technology Bureau.

Data Availability Statement: The datasets presented in this article are openly available at https://github.com/ZionG99/Wanfang-research-topic-dataset, accessed on 10 February 2025.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Behrouzi, S.; Sarmoor, Z.S.; Hajsadeghi, K.; Kavousi, K. Predicting scientific research trends based on link prediction in keyword networks. *J. Informetr.* **2020**, *14*, 101079. [CrossRef]
2. Xu, S.; Hao, L.; An, X.; Yang, G.; Wang, F. Emerging research topics detection with multiple machine learning models. *J. Informetr.* **2019**, *13*, 100983. [CrossRef]
3. Ofer, D.; Kaufman, H.; Linial, M. What's next? Forecasting scientific research trends. *Heliyon* **2024**, *10*, e23781. [CrossRef]
4. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the International Conference on Learning Representations, Virtual, 25 April 2022.
5. Zhang, X.; Xu, Y.; He, W.; Guo, W.; Cui, L. A comprehensive review of the oversmoothing in graph neural networks. In Proceedings of the CCF Conference on Computer Supported Cooperative Work and Social Computing, Harbin, China, 18–20 August 2023; pp. 451–465.
6. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
7. Taylor, S.J.; Letham, B. Forecasting at scale. *Am. Stat.* **2018**, *72*, 37–45. [CrossRef]
8. Zou, T.; Guo, P.; Li, F.; Wu, Q. Research topic identification and trend prediction of China's energy policy: A combined LDA-ARIMA approach. *Renew. Energy* **2024**, *220*, 119619. [CrossRef]
9. Yu, D.; Xiang, B. An ESTs detection research based on paper entity mapping: Combining scientific text modeling and neural prophet. *J. Informetr.* **2024**, *18*, 101551. [CrossRef]
10. Yang, Z.; Zhang, W.; Wang, Z.; Huang, X. A deep learning-based method for predicting the emerging degree of research topics using emerging index. *Scientometrics* **2024**, *129*, 4021–4042. [CrossRef]
11. Zhang, H.; Feng, L.; Wang, J.; Gao, N. Development of technology predicting based on EEMD-GRU: An empirical study of aircraft assembly technology. *Expert Syst. Appl.* **2024**, *246*, 123208. [CrossRef]
12. Ma, Q.; Fu, X.; Yang, Q.; Qiu, D. Adaptive masked network for ultra-short-term photovoltaic forecast. *Eng. Appl. Artif. Intell.* **2025**, *139*, 109555. [CrossRef]
13. Chen, Z.; Ma, M.; Li, T.; Wang, H.; Li, C. Long sequence time-series forecasting with deep learning: A survey. *Inf. Fusion* **2023**, *97*, 101819. [CrossRef]
14. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.
15. Gopali, S.; Abri, F.; Siami-Namini, S.; Namin, A.S. A comparison of TCN and LSTM models in detecting anomalies in time series data. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 2415–2420.
16. LI, Y.; Wang, Q. Adaptive genetic algorithm-optimized temporal convolutional networks for high-precision ship traffic flow prediction. *Evol. Syst.* **2025**, *16*, 1–17. [CrossRef]
17. Ye, X.; Hao, Y.; Ye, Q.; Wang, T.; Yan, X.; Chen, J. Demand forecasting of online car-hailing by exhaustively capturing the temporal dependency with TCN and Attention approaches. *IET Intell. Transp. Syst.* **2024**, *18*, 2565–2575. [CrossRef]
18. Xue, H.; Gui, X.; Wang, G.; Yang, X.; Gong, H.; Du, F. Prediction of gas drainage changes from nitrogen replacement: A study of a TCN deep learning model with integrated attention mechanism. *Fuel* **2024**, *357*, 129797. [CrossRef]
19. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3848–3858. [CrossRef]
20. Geng, H.; Wang, D.; Zhuang, F.; Ming, X.; Du, C.; Jiang, T.; Guo, H.; Liu, R. Modeling dynamic heterogeneous graph and node importance for future citation prediction. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–21 October 2022; pp. 572–581.
21. Wang, L.; Huang, Y.; Wu, H. Spatial-temporal Graph Convolutional Networks with Diversified Transformation for Dynamic Graph Representation Learning. *arXiv* **2024**, arXiv:2408.02704.
22. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
23. Fan, J.; Weng, W.; Tian, H.; Wu, H.; Zhu, F.; Wu, J. RGDAN: A random graph diffusion attention network for traffic prediction. *Neural Netw.* **2024**, *172*, 106093. [CrossRef] [PubMed]
24. Wang, L.; Zhang, Y.; Yuan, J.; Cao, S.; Zhou, B. RLGAT: Retweet prediction in social networks using representation learning and GATs. *Multimed. Tools Appl.* **2024**, *83*, 40909–40938. [CrossRef]
25. Zou, Z.; Sun, Y.; Li, W.; Li, Y.; Wang, Y. A Paper Citation Link Prediction Method Using Graph Attention Network. In Proceedings of the International Artificial Intelligence Conference, London, UK, 14–16 July 2023; pp. 32–41.
26. Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Zhang, C. Graph wavenet for deep spatial-temporal graph modeling. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 1907–1913.

27. Peng, C.; Zhang, Y. Industrial Internet of Things Intrusion Detection Model Integrating Graph Attention Network and Gated Temporal Convolutional Network. In Proceedings of the 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE), Shanghai, China, 1–3 March 2024; pp. 596–599.
28. Vaswani, A. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017.
29. Doostmohammadian, M.; Qureshi, M.I.; Khalesi, M.H.; Rabiee, H.R.; Khan, U.A. Log-Scale Quantization in Distributed First-Order Methods: Gradient-based Learning from Distributed Data. *IEEE Trans. Autom. Sci. Eng.* **2025**. [CrossRef]