

Topology-Aware Dynamic Power Flow Calculation Method for Power Grid Based on Graphormer-GE

Huiqin Li¹, Xingbo Zhang¹

¹School of Information Science and Engineering, Shandong Agriculture and Engineering University, Jinan 250100, Shandong, China

Corresponding author: Xingbo Zhang (e-mail: 13356694319@163.com)

ABSTRACT To balance prediction accuracy and computational efficiency in grid modeling, this paper proposes Graphormer-GE, a topology-aware dynamic power flow calculation method. It not only reflects the steady-state power and voltage distribution at a certain moment but also characterizes the continuous evolution behavior of the system under the influence of time-varying factors such as topology disturbances, load fluctuations, and new energy integration. The proposed Graphormer-GE method addresses the dynamic power flow challenge by constructing a high-dimensional graph integrating electrical characteristics and real-time communication status from IoT devices. It embeds electrical distance and physical topology through a geometric encoding mechanism to enhance sensitivity to topology changes. By establishing a unified attention computation space and jointly modeling temporal evolution and spatial relationships among nodes, the expressive capability of the model is significantly improved. Furthermore, a dynamic graph update strategy incorporating communication link quality ensures stable performance in network-constrained environments. The proposed framework provides an effective computational paradigm for intelligent power systems and electromagnetic energy transmission networks requiring accurate state perception and propagation analysis. Experiments on the IEEE 300-node system demonstrate superior accuracy (MAE 0.037 p.u.) and response speed (Postfault Error 0.07 p.u.), while maintaining stable performance during topological disturbances (error within 0.05 p.u.) across diverse communication environments, confirming its practical value for distributed power system monitoring and reliable electromagnetic infrastructure operation.

INDEX TERMS dynamic power flow calculation, Graphormer-GE, grid topology awareness, electromagnetic energy transmission, geometric encoding

I. INTRODUCTION

WITH the increasing penetration of new energy sources in modern power systems and the enhanced flexibility of the load side, the operation status of the power grid has become highly dynamic and uncertain, creating challenges not only for system stability but specifically for energy-intensive industrial consumers. For the industry, which relies on a stable, high-quality power supply to prevent machinery damage and maintain consistent product quality, this heightened dynamism necessitates advanced and real-time modeling solutions [1-3]. Traditional static power flow calculation based on the Newton-Raphson method is difficult to meet the real-time perception and prediction needs of the dynamic behavior of complex power grids [4-6]. In recent years, graph neural networks have been widely used in the field of power flow prediction due to their advantages in modeling power grid topology [7-9]. However, existing methods often have difficulty in achieving good coordination between prediction accuracy

and computational efficiency when facing large-scale, high-dimensional and time-varying power grid scenarios, which limits their practical application in online analysis and edge deployment.

This paper focuses on the dynamic power flow modeling problem under the condition of frequent changes in power grid topology and limited communication environment. This type of problem widely exists in the operation of modern smart grids, especially in typical scenarios such as distributed energy access, N-1 fault recovery, and load mutation. Ma R [10] found that for AC and DC distribution networks with high penetration of distributed photovoltaic access, when the flexible interconnection device fails, the photovoltaic access area can cause the voltage to gradually increase and the frequency to fluctuate violently or even become unstable. Mehmood M Y [11] proposed that the introduction of edge computing platform improves the efficiency of model deployment. Althoubi A [12] said that the data delay collected by IoT devices can cause

deviations in model input, thereby affecting the prediction results. Feng X [13] proposed a dynamic backoff method for multi-node communication of AUV (Autonomous Underwater Vehicle) based on Carrier Sense Multiple Access with Collision Avoidance, CSMA/CA, to realize dynamic multi-node communication in underwater electric fields. Alam M J E [14] pointed out that solar and wind power output can cause rapid voltage/power changes. Therefore, how to improve the model's ability to perceive topology changes and network adaptability under complex operating conditions has become a key issue that needs to be solved urgently. The paper introduces Graphormer-GE, a topology-aware dynamic power flow modeling framework designed to address stability challenges in modern grids. The architectural foundations for enhanced grid state perception stem from decentralized graph neural networks and two-stream spatio-temporal modeling approaches [15,16]. This method constructs a high-dimensional dynamic graph that fuses electrical characteristics (e.g., voltage, power) with real-time communication status (e.g., delay). A geometric encoding mechanism calculates electrical distance via the Floyd-Warshall algorithm to enhance the model's sensitivity to structural changes. The core innovation lies in a hybrid spatiotemporal attention mechanism that uses 1D- CNN and GRU for temporal feature extraction and multi-head graph attention for spatial modeling, effectively capturing the complex spatio-temporal coupling in the grid state. Hierarchical decomposition within self-attention structures improves the reliability of modeling long-term energy load fluctuations [17]. Crucially, a dynamic graph update strategy is designed for network-grid collaboration, embedding communication quality constraints into attention weights. The stability of the system under fluctuating network conditions is reinforced by lightweight graph convolutional accelerators and interactive balance strategies between distribution and mobile networks [18,19]. This directly benefits the high-reliability demands of the industry by maintaining stable power quality monitoring despite communication limitations, thus enhancing the model's robustness. The framework is trained using a multi-task loss function and achieves strong performance, keeping prediction error below 0.05 p.u. during major disturbances, validating its practical value for distributed industrial control.

II. ALGORITHM DESIGN

POWER GRID GRAPH MODELING BASED ON GRAPHORMER-GE

This paper first constructs a high-dimensional dynamic graph structure that integrates electrical characteristics and communication status information, and designs a graph neural network model input form with topology awareness based on the Graphormer architecture. The overall architecture is shown in Figure 1 [20- 22]. The power grid system can be abstracted as a directed graph:

$$G = (V, E) \quad (1)$$

The node set $V = \{v_1, v_2, \dots, v_N\}$ represents the busbar in the power grid, the edge set $E \subseteq V \times V$ represents the branch connection relationship, and each edge $e_{ij} \in E$ describes the power transmission path from node v_i to v_j . In the dynamic modeling process, the power grid topology may change due to line tripping, load changes, or renewable integration. The time dimension needs to be introduced and the dynamic graph structure is defined as:

$$G_t = (V, E_t, X_t, A_t) \quad (2)$$

$t \in T$ represents the discrete time step, $T = \{1, 2, \dots, T\}$; E_t is the edge set at time t ; $X_t \in \mathbb{R}^{N \times d_v}$ is the node feature matrix; $A_t \in \{0, 1\}^{N \times N}$ is the adjacency matrix, which represents the grid connection relationship at the current moment. The grid node features include physical quantities such as voltage amplitude $|V_i|$, phase angle θ_i , active power P_i , reactive power Q_i , etc. Combined with the data collected in real time by IoT (Internet of Things) sensors, the node feature vector is defined as:

$$x_{i,t} = [|V_{i,t}|, \theta_{i,t}, P_{i,t}, Q_{i,t}, \dot{P}_{i,t}, \dot{Q}_{i,t}, S_{i,t}, f_{i,t}] \quad (3)$$

$\dot{P}_{i,t}$ is the active power injection of node i at time t , $\dot{Q}_{i,t}$ is the reactive power injection of node i at time t , $S_{i,t}$ is the apparent power, and $f_{i,t}$ is the frequency measurement value. This extended feature set enhances the model's ability to respond to dynamic disturbances. The branch parameters include admittance Y_{ij} , impedance Z_{ij} , line capacity C_{ij} , and current load rate $L_{ij,t}$. The edge feature vector is defined as:

$$e_{ij,t} = [\text{Re}(Y_{ij}), \text{Im}(Y_{ij}), |Z_{ij}|, C_{ij}, L_{ij,t}, \tau_{ij,t}, b_{ij,t}]^T \quad (4)$$

$\tau_{ij,t}$ is the communication link delay, $b_{ij,t}$ is the available bandwidth, and these communication quality indicators can be used in the subsequent dynamic update mechanism of the graph structure. Since the power grid topology changes with the operating state, the adjacency matrix A_t of the graph should be dynamically adjusted according to the actual connection relationship and communication link status. The graph evolution function is defined as follows:

$$A_t = F_G(A_{t-1}, \Delta G_t) \quad (5)$$

ΔG_t represents the graph structure change at time t relative to the previous period, including the addition/disconnection of branches, node status changes, etc. Function $F_G(\cdot)$ can be obtained in real time through the graph database or SCADA (Supervisory Control and Data Acquisition) system. In order to enhance the model's sensitivity to the grid topology, the geometry encoding (GE) mechanism is introduced in the graph modeling stage, and the electrical distance between nodes in the grid is used as the position encoding information to embed the model input. Assume that the shortest path length matrix between nodes

is $D \in \mathbb{R}^{N \times N}$, and its element D_{ij} represents the electrical distance from node i to node j , then the geometric encoding vector is defined as:

$$PE_i = \phi(D_{i,:}) = \frac{1}{N} \sum_{j=1}^N \text{ReLU}(W_{pe} \cdot [D_{ij}, \|v_i - v_j\|_2]) \quad (6)$$

$W_{pe} \in \mathbb{R}^{d_{pe} \times 2}$ is a learnable parameter matrix, and $\phi(\cdot)$ is a nonlinear mapping function used to convert electrical distance into position encoding in high-dimensional space. Finally, the geometric encoding is concatenated with the original node features to form an enhanced input representation:

$$\tilde{x}_{i,t} = \text{LayerNorm}(x_{i,t} \oplus PE_i) \quad (7)$$

$\oplus$ represents the vector concatenation operation, and LayerNorm is the layer normalization function to improve the stability of model training. The power grid graph structure constructed in this section integrates the physical characteristics of the power grid and the operation data, and also introduces a geometric encoding mechanism to improve the model's perception of topological changes, laying a solid foundation for the subsequent hybrid spatiotemporal attention modeling based on Graphormer.

Figure 1 shows the overall process of the grid topology-aware dynamic power flow calculation method based on Graphormer-GE: starting from the physical system of the grid, the operation data of nodes and branches and communication status information are collected in real time through IoT sensors. Then a dynamic graph structure that integrates electrical characteristics and communication quality is constructed, and a geometric encoding mechanism is introduced to enhance the model's ability to perceive topology changes. Then the enhanced input is fed into the Graphormer-GE model, and the time evolution and spatial propagation characteristics of the power grid state are extracted through hybrid spatiotemporal attention modeling; combined with the dynamic graph update strategy of network-grid collaboration, the graph structure of the next moment is predicted based on the current state; finally, the dynamic flow results such as voltage, phase angle, and power are output to achieve accurate prediction and rapid response to the power grid operation state.

TOPOLOGY ADAPTIVE GEOMETRIC CODING MECHANISM

This paper proposes a topology-adaptive GE mechanism that introduces the electrical distance between nodes and the physical topology information to enhance the modeling stability and prediction accuracy under topological disturbances [23]. The power grid diagram is $G = (V, E)$, and the electrical distance between nodes i and j is defined as:

$$d_{ij}^{\text{elec}} = \sum_{e \in P_{ij}} z_e \quad (8)$$

P_{ij} represents the shortest path from node i to node j , and z_e is the impedance value of branch e . This electrical distance reflects the actual path cost of energy transmission in the power grid. In order to efficiently obtain the global electrical distance matrix, the Floyd-Warshall algorithm is used to solve the shortest path length between any two nodes [24,25], and the electrical distance matrix $D \in \mathbb{R}^{N \times N}$ is constructed, whose element D_{ij} represents the electrical distance from node i to node j , satisfying the following recursive relationship:

$$D_{ij}^{(k)} = \min(D_{ij}^{(k-1)}, D_{ik}^{(k-1)} + D_{kj}^{(k-1)}) \quad (9)$$

$D_{ij}^{(0)} = z_{ij}$, if there is no direct connection between nodes, it takes infinity; after $k = 1$ to N iterations, the complete electrical distance matrix D is finally obtained. In order to further improve the sensitivity of the model to topological changes, the electrical distance matrix D is mapped to the node-level geometric encoding vector $PE_i \in \mathbb{R}^{d_{pe}}$. After normalizing each row $D_{i,:} \in \mathbb{R}^N$, the learnable mapping function $\Phi: \mathbb{R}^N \rightarrow \mathbb{R}^{d_{pe}}$ is obtained:

$$PE_i = \Phi(D_{i,:}) = W_{pe} \cdot \sigma\left(\frac{D_{i,:} - \mu_D}{\sigma_D}\right) \quad (10)$$

$$\mu_D = \frac{1}{N^2} \sum_{i,j} D_{ij}, \quad \sigma_D = \sqrt{\frac{1}{N^2} \sum_{i,j} (D_{ij} - \mu_D)^2}$$

represents the mean and standard deviation of the electrical distance matrix; $\sigma(\cdot)$ is the nonlinear activation function; $W_{pe} \in \mathbb{R}^{d_{pe} \times N}$ is the learnable parameter matrix. Considering that the topology may change frequently during the operation of the power grid, a topology adaptive mechanism is designed to enable the geometric coding to be dynamically adjusted with the power grid state. The graph evolution function ΔG_t is defined to represent the topology change at the t th moment relative to the previous period, and the updated geometric coding is:

$$PE_i^{(t)} = \gamma \cdot PE_i^{(t-1)} + (1 - \gamma) \cdot \Phi(D_t) \quad (11)$$

$\gamma \in [0, 1]$ is the attenuation coefficient, which controls the weight distribution of historical coding and current topology information, ensuring that the model still has stable modeling capabilities when the topology changes frequently. The topology adaptive geometric coding mechanism proposed in this section achieves highly sensitive modeling of power grid structure changes by integrating electrical distance and physical topology information. The ablation experiment results of the topology adaptive geometric coding mechanism in the IEEE 300-node system are shown in Table 1.

HYBRID SPATIOTEMPORAL ATTENTION MECHANISM

This paper proposes a hybrid spatio-temporal attention mechanism, which integrates local time series modeling and graph structure spatial attention mechanism to improve the

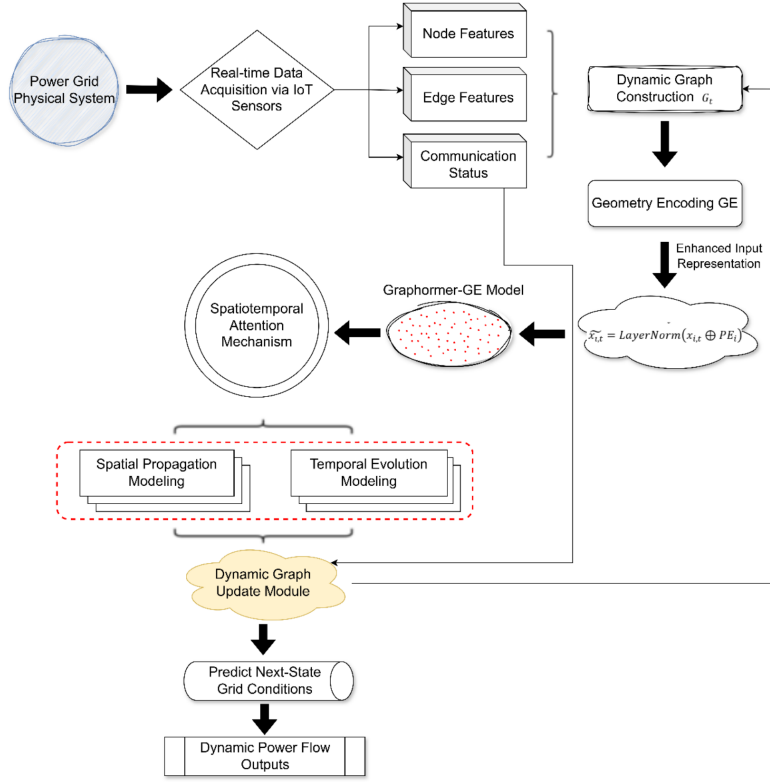


FIGURE 1. Overall architecture diagram

TABLE 1. Ablation experiment

Model Variant	MAE (p.u.)	Post-fault Error (p.u.)
Physical Topology Only	0.058	0.14
Electrical Distance Only	0.051	0.11
Full Model	0.037	0.07

prediction accuracy and response speed of the model for the dynamic power flow distribution of the power grid [26-28]. Considering that the operation status of the power grid has obvious time series evolution characteristics, such as load fluctuations, changes in new energy output, etc., this paper first uses a one-dimensional convolutional network to extract the local time series pattern of the node state:

$$h_i^{(c)} = \text{ReLU}(W_c * x_{i,t-L+1:t} + b_c) \quad (12)$$

$x_{i,t-L+1:t} \in \mathbb{R}^{L \times d}$ represents the historical feature sequence of node i in the time window $[t-L+1, t]$,

$W_c \in \mathbb{R}^{k \times d}$ is the convolution kernel parameter, k is the filter size, L is the time window length, and $h_i^{(c)} \in \mathbb{R}^{d'}$ is the extracted local time series feature. In order to further capture long-term dependencies, a gated recurrent unit is introduced to model the time series of the convolution output:

$$r_t = \sigma(W_r \cdot [h_t^{(c)}, h_{t-1}^{(g)}]) \quad (13)$$

$$z_t = \sigma(W_z \cdot [h_t^{(c)}, h_{t-1}^{(g)}]) \quad (14)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_t^{(c)}, h_{t-1}^{(g)}]) \quad (15)$$

$$h_t^{(g)} = (1 - z_t) \odot h_{t-1}^{(g)} + z_t \odot \tilde{h}_t \quad (16)$$

r_t and z_t are reset gate and update gate respectively, $\sigma(\cdot)$ is Sigmoid function, $\odot$ represents element-by-element multiplication, and $h_t^{(g)} \in \mathbb{R}^{d'}$ is the final time dimension encoding representation. In terms of spatial dimension modeling, this paper designs a multi-head graph attention mechanism (MGA) to capture the complex electrical interaction relationship between nodes in the power grid [29,30]. For any node i , the attention weight between it and its neighbor node $j \in \mathcal{N}(i)$ is defined as:

$$\alpha_{ij}^{(k)} = \text{Softmax}_j \left(\frac{(W_q^{(k)} h_i)^T (W_k^{(k)} h_j)}{\sqrt{d_k}} \right) \quad (17)$$

$W_q^{(k)}, W_k^{(k)} \in \mathbb{R}^{d' \times d_k}$ are the learnable parameter matrices in the k th attention head, d_k is the scaling factor, and the Softmax operation is performed on the neighbor set $\mathcal{N}(i)$ of node i .

$$\text{Head}_k(h_i) = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(k)} W_v^{(k)} h_j \quad (18)$$

After the multi-head mechanism is introduced, the outputs of each head are concatenated and mapped to the original dimension to form the final spatial attention representation:

$$h_i^{(m)} = \text{Concat}(\text{Head}_1(h_i), \dots, \text{Head}_H(h_i)) W^O \quad (19)$$

$W^O \in \mathbb{R}^{Hd_k \rightarrow d'}$ is the output projection matrix. This mechanism effectively enhances the model's ability to

model non-local dependencies in the power grid topology. The time dimension encoding $h_i^{(g)}$ is fused with the space dimension encoding $h_i^{(m)}$ to construct a unified spatiotemporal embedding representation:

$$h_i^{(st)} = \text{LayerNorm}(h_i^{(g)} + h_i^{(m)}) \quad (20)$$

This fusion strategy enables the model to capture the dynamic evolution trend of the power grid state and reflect its propagation path in the spatial topology, significantly improving the power flow prediction capability under complex operating conditions.

DYNAMIC GRAPH UPDATE MECHANISM FOR NETWORK-GRID COLLABORATION

This paper proposes a network-power coordinated graph update mechanism. The communication link quality information can be embedded in the adjacency matrix of the graph neural network to achieve joint modeling of the power system and the communication network [31,32]. Assume that there is a physical electrical connection between the power grid nodes $i, j \in \mathcal{V}$, that is, the branch $(i, j) \in \mathcal{E}$. If the corresponding communication link is used to transmit state information, the state vector of the link is defined as:

$$S_{ij} = [\tau_{ij}, b_{ij}, l_{ij}] \quad (21)$$

$\tau_{ij} \in \mathbb{R}^+$ represents the communication delay from node i to node j ; $b_{ij} \in \mathbb{R}^+$ represents the available bandwidth; $l_{ij} \in [0, 1]$ represents the packet loss rate. A communication state-aware adjacency weight decay function $f : S_{ij} \rightarrow (0, 1]$ is designed so that the adjacency matrix elements are no longer binary, but are dynamically adjusted according to the quality of the communication link:

$$a'_{ij} = a_{ij} \cdot f(S_{ij}) = a_{ij} \cdot \sigma \left(w_\tau \cdot \frac{1}{\tau_{ij}} + w_b \cdot b_{ij} - w_l \cdot l_{ij} \right) \quad (22)$$

$\sigma(\cdot)$ is a Sigmoid function used to map the comprehensive score to the interval $(0, 1]$; $w_\tau, w_b, w_l \in \mathbb{R}^+$ are learnable weight coefficients that controls the influence of each communication indicator on the graph structure; if there is no direct communication link or the link is unavailable, $f(S_{ij}) = 0$, the corresponding edge weight is set to zero. The time decay factor $\beta \in [0, 1]$ is further introduced to

balance the influence of the historical communication status and the current link quality:

$$f_t(S_{ij}) = \beta \cdot f_{t-1}(S_{ij}) + (1 - \beta) \cdot f(S_{ij}^{(t)}) \quad (23)$$

This mechanism enables the graph structure to evolve dynamically with changes in the communication network status, avoiding information dissemination failures caused by sudden network congestion or link interruption. The quality of the communication link is further introduced into the attention module to improve the robustness of the model in network-constrained scenarios. In the original graph attention mechanism, the attention weight of node i to neighbor j is formula (17). After introducing the communication status, the attention weight is weighted and modified:

$$\alpha'_{ij} = \frac{\alpha_{ij} \cdot f(S_{ij})}{\sum_{k \in \mathcal{N}(i)} \alpha_{ik} \cdot f(S_{ik})} \quad (24)$$

This formula ensures that when the communication quality is poor, the attention weights between nodes are automatically attenuated to prevent misprediction due to information delay or loss. The aggregate feature of node i is updated as:

$$h'_i = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha'_{ij} W_v h_j \right) \quad (25)$$

W_v is a learnable parameter matrix, and $\sigma(\cdot)$ is a non-linear activation function. In order to enhance the model's adaptability to communication link fluctuations, a communication quality constraint term is introduced into the loss function:

$$L_{\text{QoS}} = \lambda \cdot \sum_{i,j} \left| \log \left(\frac{f(S_{ij})}{\hat{f}(S_{ij})} \right) \right| \quad (26)$$

$\hat{f}(S_{ij})$ is the communication impact factor estimated by the model, and λ is a hyperparameter used to balance the weight between this constraint and other task objectives, as shown in Figure 2. This strategy is a hybrid dynamic update mechanism based on minute-level sampling and capable of responding to key communication events. By directly embedding communication quality information into the graph's topology and attention weights, it enables the model to adaptively adjust information propagation paths in network-constrained environments, thereby maintaining stable prediction performance.

Figure 2 shows the dynamic graph update mechanism process of network-grid collaboration. Starting from the physical topology of the grid, the communication link quality information of nodes and branches is collected to construct the communication state vector. This information is used in the dynamic weighted adjacency matrix to convert the original binary connection relationship into a continuous weight reflecting the communication quality. The updated

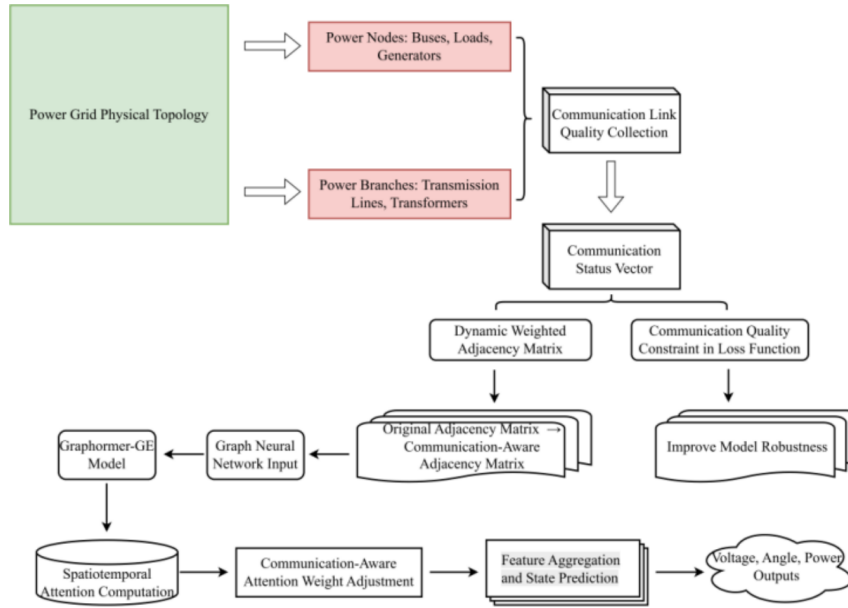


FIGURE 2. Schematic diagram of network-grid joint topology

graph structure is used as the input of the graph neural network and enters the Graphormer-GE model for hybrid spatiotemporal attention calculation. In this process, the communication state further participates in the adjustment of the attention weight, so that the model automatically reduces the influence weight between corresponding nodes when the communication quality decreases. At the same time, the communication link quality information is also introduced into the loss function as an additional constraint term to further improve the prediction robustness of the model in a network-constrained environment. The final output voltage, phase angle, power and other flow results can achieve accurate modeling and stable prediction of the power grid state under complex operating environments.

MODEL TRAINING AND OPTIMIZATION FRAMEWORK

This paper adopts a multi-task learning framework to jointly model multiple key state variables such as bus voltage amplitude, phase angle and branch active power, and introduces topological disturbance recovery error and communication quality constraints to construct a comprehensive optimization objective function [33-35]. To adapt the Graphormer architecture for this continuous regression task of dynamic power flow prediction, the original classification head is removed. A linear regression head is applied directly to the hidden state of each node from the final Graphormer encoder layer to output continuous values. Assume that the model input is the power grid graph structure $G_t = (V, E_t, X_t, A_t)$ at time step t , and the output is the target state variable set:

$$\hat{Y}_t = [\hat{V}_t, \hat{\theta}_t, \hat{P}_t] \quad (27)$$

$\hat{V}_t \in \mathbb{R}^N$ predicts the bus voltage amplitude; $\hat{\theta}_t \in \mathbb{R}^N$ predicts the bus voltage phase angle; $\hat{P}_t \in \mathbb{R}^M$ predicts the branch active power, M is the number of branches at the

current moment. The corresponding true label is recorded as $Y_t = [V_t, \theta_t, P_t]$. The model uses a weighted multi-task loss function for end-to-end training:

$$L_{\text{total}} = \lambda_1 L_{\text{MAE}} + \lambda_2 L_{\text{Post-fault}} + \lambda_3 L_{\text{QoS}} \quad (28)$$

The definitions of each part are as follows:
Main task loss term (mean absolute error):

$$L_{\text{MAE}} = \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{N} \|\hat{V}_t - V_t\|_1 + \frac{1}{N} \|\hat{\theta}_t - \theta_t\|_1 + \frac{1}{M} \|\hat{P}_t - P_t\|_1 \right) \quad (29)$$

This loss term is used to measure the prediction accuracy of the model for various physical quantities under the steady-state operation of the power grid.

The topological disturbance recovery error term introduces additional penalty terms after line tripping, renewable integration, or load mutation events to enhance the model's adaptability to topological changes:

$$L_{\text{Post-fault}} = \frac{1}{T_f} \sum_{t \in T_f} (\|\hat{V}_t - V_t\|_2^2 + \|\hat{\theta}_t - \theta_t\|_2^2 + \|\hat{P}_t - P_t\|_2^2) \quad (30)$$

$T_f \subset \{1, \dots, T\}$ represents the set of time points where topological disturbance occurs, and $T_f = |T_f|$ is the number of disturbances.

In order to enable the model to maintain stable performance in a network-constrained environment, the correlation loss between the communication link quality and the attention weight is further introduced:

$$L_{\text{QoS}} = \frac{1}{T} \sum_{t=1}^T \sum_{i,j} \left| \log \left(\frac{f(S_{ij}^{(t)})}{\hat{f}(S_{ij}^{(t)})} \right) \right| \quad (31)$$

$f(S_{ij}^{(t)})$ is the actual communication impact factor between nodes i and j at time t ; $\hat{f}(S_{ij}^{(t)})$ is the communication impact factor estimated by the model. This loss term encourages the model to learn to reasonably reflect the quality of the communication link in the attention weight. The model is trained using the Adam optimizer and dynamically adjusts the learning rate in combination with the cosine annealing scheduler:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_0 - \eta_{\min}) \cdot \left[1 + \cos \left(\frac{\pi t}{T_{\max}} \right) \right] \quad (32)$$

η_t is the learning rate of the t th round; $\eta_{\min}$ is the minimum learning rate; $T_{\max}$ is the total number of training rounds. To prevent gradient explosion, the gradient clipping strategy is adopted, and the maximum gradient norm threshold $g_{\max}$ is set, that is:

$$g = \nabla L_{\text{total}}, \quad \tilde{g} = \min \left(1, \frac{g_{\max}}{\|g\|_2} \right) \cdot g \quad (33)$$

Finally, the updated gradient $\tilde{g}$ is used to update the parameters. The model parameters are shown in Table 2.

III. EXPERIMENT AND VERIFICATION

EXPERIMENTAL DESIGN

This experiment selects IEEE standard systems and actual power grid operation data sets for model training and verification. All systems are equipped with SCADA acquisition devices and IoT (Internet of Things) edge nodes to achieve minute-level real-time acquisition of key variables such as bus voltage, node phase angle, and regional load power. To ensure the fairness and validity of the experimental comparison, this paper adopts a unified hybrid architecture framework when evaluating the performance of GCN, GAT and GraphSAGE graph neural networks: all models combine the same temporal modeling module to capture the temporal evolution characteristics of the power grid state.

Figure 3 shows the local power grid topology and IoT sensor deployment diagram, including some node and branch connection relationships. Nodes represent buses, and edges represent branch connections. Some key nodes are marked as IoT data collection equipment deployment points, which are used to achieve minutelevel real-time perception of the power grid operation status. Some nodes in the figure are not directly connected to other nodes, which reflects the locality and dynamics of the power grid topology, and also reflects the deployment flexibility of the proposed method in edge access and network-constrained scenarios.

EVALUATION OF POWER FLOW PREDICTION ACCURACY

For the evaluation of power flow prediction accuracy, the mean absolute error (MAE), root mean square error (RMSE), determination coefficient R2, maximum error (Max Error) and post-fault error are used as evaluation indicators. Post-fault error is defined as the mean absolute error within 3.5 seconds after the occurrence of fault N-1.

Figure 4 shows the performance of four methods (GCN, GAT, GraphSAGE, Graphormer-GE) in power flow prediction tasks under different grid scales (IEEE 39, 118 and 300 node systems). The evaluation indicators include MAE, RMSE, R2, Max Error and Post-fault Error. The horizontal axis represents different model methods, and the vertical axis represents the corresponding indicator values. In Figure 4(a), the MAE of Graphormer-GE is 0.028, which is significantly lower than 0.045 of GCN, 0.038 of GAT, and 0.041 of GraphSAGE; the RMSE is only 0.42, which is much better than other methods. As the scale of the power grid increases to 118 and 300 nodes, Graphormer-GE has obvious advantages. In the IEEE 300-node system, the MAE is 0.037, the RMSE is 0.55, and its Post-fault Error remains at 0.07, while other methods generally exceed 0.13. This shows that Graphormer-GE not only performs well in small-scale systems, but also has stronger generalization and robustness in large-scale complex power grids.

By introducing a geometric encoding mechanism, electrical distance and physical topology information are embedded into the model input, making the model more sensitive and adaptable to topology changes. The hybrid spatiotemporal attention mechanism effectively captures the coupling characteristics of the grid state in time evolution and spatial propagation, improving the overall modeling accuracy. The dynamic graph update strategy combines the communication link quality to optimize the graph structure, so that the model can still maintain stable prediction performance in edge deployment and network-constrained environments. These designs work together to enable Graphormer-GE to achieve better prediction accuracy and response speed in multi-scale power grid systems, especially in the recovery phase after topological disturbances, showing faster convergence speed and lower error accumulation.

EVALUATION OF TOPOLOGICAL CHANGE RESPONSE CAPABILITY

In order to evaluate the model's responsiveness to topological changes, an N-1 fault simulation test is designed. The error accumulation response curve is used to measure the speed and stability of the model's recovery from topological disturbances.

Figure 5 shows the prediction response curves of different methods for the grid bus voltage under N-1 fault disturbance. The horizontal axis represents time, reflecting the time process after the fault occurs; the vertical axis represents the voltage amplitude, reflecting the change trend of the key bus voltage. The measured voltage curve is used as a

TABLE 2. Model parameter setting table

Parameter Category	Parameter Name	Value / Description
Model Architecture	Node Embedding Dimension	128
	Edge Feature Dimension	16
	Number of Attention Heads	4
	Feed-Forward Network Hidden Dimension	512
	Number of Graphormer Layers	4
	Geometry Encoding Dimension	16
	Use Layer Normalization	Yes
Temporal Modeling Module	Use Dropout	Yes
	Convolutional Kernel Size (1D-CNN)	3
	GRU Hidden State Dimension	128
Loss Function Setup	Main Task Loss	Mean Absolute Error (MAE)
	Topology Perturbation Loss Weight	0.5
	Communication Quality Loss Weight	0.3
	Multi-task Loss Combination	Weighted Sum
Training Configuration	Optimizer	Adam
	Initial Learning Rate	5e-4
	Minimum Learning Rate	1e-5
	Learning Rate Scheduler	Cosine Annealing
	Total Training Epochs	200
	Batch Size	32
	Gradient Clipping Threshold	1
	Early Stopping	Yes

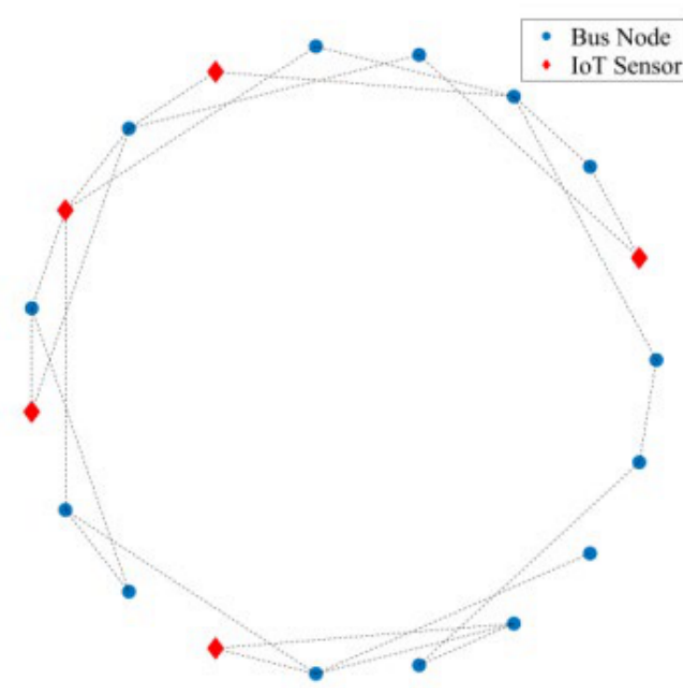


FIGURE 3. Schematic diagram of IoT data collection equipment deployment

benchmark to show the dynamic change process of voltage over time in the actual system. As can be seen from Figure 5, there are differences in the response speed and accuracy of each method after the fault occurs. For example, at $t = 0.5$ seconds, the measured voltage drops to 0.94 p.u., while the predicted values of GCN are 0.96 p.u., GAT is 0.97 p.u., GraphSAGE is 0.96 p.u., and Graphormer-GE is 0.93 p.u. As time goes by, the predicted values of Graphormer-

GE approach the measured voltage. Compared with the predicted values of 0.92 p.u. by GCN and GAT at $t = 2$ seconds, the error is large. Graphormer-GE has completely converged to the measured voltage (0.9 p.u.) at $t = 3.5$ seconds, while other methods still have certain deviations, with the predicted values of 0.92 p.u. by GCN and 0.91 p.u. by GAT. This shows that Graphormer-GE responds faster and has higher prediction accuracy under topological pertur-

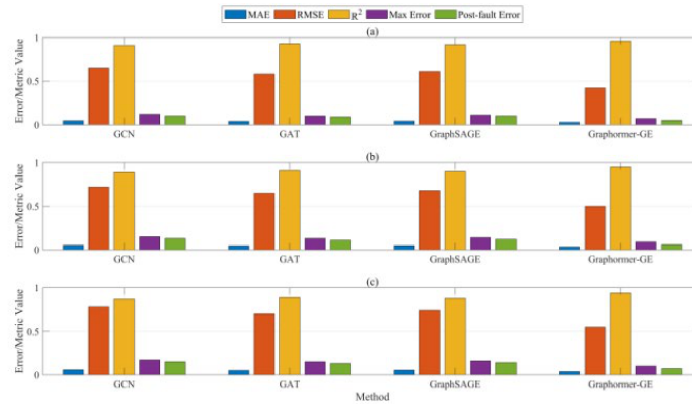


FIGURE 4. Evaluation of power flow prediction accuracy. Figure 4(a) IEEE 39-bus System; Figure 4(b) IEEE 118-bus System; Figure 4(c) IEEE 300-bus System

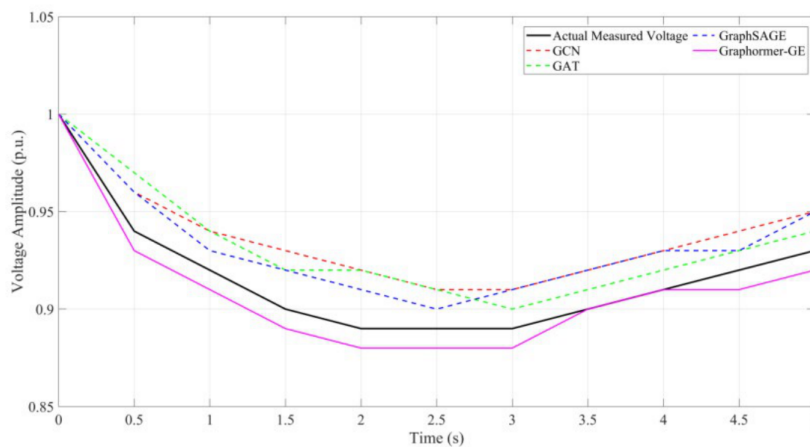


FIGURE 5. Comparison of voltage prediction response curves under topological disturbance

bations. From the response curve in Figure 5, Graphormer-GE shows a relatively smooth transition behavior during the recovery process after the topological disturbance, without obvious oscillation or error repetition. The dynamic graph update mechanism can adjust the graph structure according to the real-time communication link status, so that the model can still maintain stable prediction performance in edge deployment and network-constrained environments. Thanks to the modeling ability of the geometric coding strategy for electrical distance, Graphormer-GE can quickly reconstruct the correlation strength between nodes after the topological mutation, achieving more accurate voltage distribution prediction. Graphormer-GE has higher prediction accuracy and shows stronger engineering practical value in terms of dynamic adaptability and robustness, which is suitable for promotion and application in complex and changeable smart grid environments.

COMPUTATIONAL EFFICIENCY AND SCALABILITY EVALUATION

To verify the applicability of the model in large-scale systems, the inference time, parameter quantity and GPU

memory usage were tested under IEEE 118 nodes, 300 nodes and 2000 nodes of the power grid.

Figure 6(a) is a comparison of inference time; Figure 6(b) is a comparison of model parameter quantity; Figure 6(c) is a comparison of GPU memory usage. Figure 6 shows the inference time, model parameter quantity and GPU memory usage of GCN, GAT, GraphSAGE and Graphormer-GE in three system scales of IEEE 118, 300 and power grid 2000 nodes. The horizontal axis represents different methods, and the vertical axis represents time, parameter amount, and memory, respectively. As can be seen from Figure 6(a), in the IEEE 118-node system, the reasoning time of Graphormer-GE is 12 ms, slightly lower than the 15 ms of GCN. As the system scale expands to 2000 nodes, the reasoning time of Graphormer-GE increases to 50 ms, slightly higher than the 40 ms of GCN, and significantly better than GAT (75 ms) and GraphSAGE (60 ms). In terms of parameter amount, Graphormer-GE is 0.8 M, 1.1 M, and 2.5 M, respectively, which are all higher than GCN (0.5 M, 0.7 M, and 1.2 M), and lower than GAT and GraphSAGE. The GPU memory usage trend is similar. Graphormer-GE is 3.0 GB in a 2000-node system, which is higher than GCN's

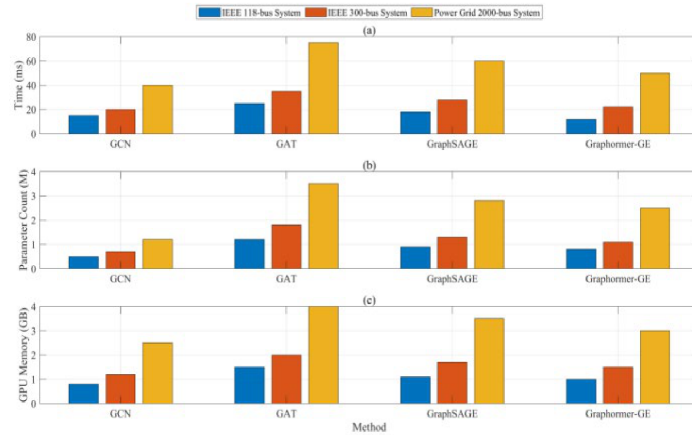


FIGURE 6. Computational efficiency and scalability evaluation

2.5 GB, but still has good deployment feasibility.

From the perspective of model structure, Graphormer-GE has a slight increase in inference time, parameter volume, and GPU memory usage compared to GCN, mainly due to its introduced geometric encoding mechanism and hybrid spatiotemporal attention structure. These modules significantly enhance the granularity of modeling the relationship between nodes by introducing electrical distance information and spatial propagation weights, but also bring higher matrix operation complexity. In the IEEE 300-node system, the reasoning time of Graphormer-GE is 10% longer than that of GCN, reflecting its more complex feature fusion process. The dynamic graph update strategy can adjust the graph structure in real time according to the status of the communication link. Although this mechanism improves the robustness of the model in a network-constrained environment, it also increases the cost of graph reconstruction in the forward propagation. Although Graphormer-GE is slightly higher than GCN in terms of computing resources, its enhanced topology expression capability and communication quality adaptation mechanism make it more suitable for deployment in actual power grid environments with heterogeneous network conditions.

MULTI-SCENARIO GENERALIZATION CAPABILITY EVALUATION

The model robustness is tested under various operating conditions, including high load conditions, low-frequency oscillations, and distributed generation (DG) fluctuations.

Figure 7 shows the voltage prediction error change trends of GCN, GAT, GraphSAGE and Graphormer-GE in 10 operating scenarios. The horizontal axis represents different grid operating conditions, including high load, low-frequency oscillation, DG fluctuation, communication delay, equipment aging, topology change, harmonic interference, load mutation, renewable integration and seasonal variation, and the vertical axis represents the prediction error. The error of GCN in the topological disturbance scenario is 0.09 p.u., and the error rises to 0.12 p.u. in the harmonic interference sce-

nario, showing strong sensitivity to complex environments. The errors of GAT and GraphSAGE are slightly lower than GCN overall. Compared with Graphormer-GE, the errors in all scenarios remain at a low level, such as only 0.05 p.u. in the equipment aging scenario and 0.04 p.u. in the renewable integration scenario, and the error fluctuation range is the smallest, showing stronger stability and adaptability.

From the error trend, the modeling performance of traditional graph neural networks is obviously limited when facing changes in node connection relationships in the power grid or decreased communication quality. For example, in the scenarios of communication delay and topological disturbance, the errors of GCN and GAT reached 0.08 p.u., 0.09 p.u. and 0.07 p.u., 0.08 p.u., respectively, reflecting their strong dependence on fixed graph structures and difficulty in dynamically adjusting the information transmission path between nodes. Graphormer-GE embeds electrical distance information into the model input by introducing a geometric encoding mechanism, and the error is 0.04 p.u. in the topological perturbation scenario, which is better than other methods. The dynamic graph update strategy designed by it can optimize the graph attention weight according to the real-time communication link status, and the error is controlled within 0.04 p.u. in the communication delay scenario, reflecting stronger edge deployment adaptability. These mechanisms work together to enable Graphormer-GE to maintain a high prediction stability under various complex operating conditions.

EDGE DEPLOYMENT PERFORMANCE EVALUATION IN MULTIPLE NETWORK ENVIRONMENTS

To verify the stability of the model in network-restricted scenarios, edge inference services were deployed in LAN, 5G network, and public Internet environments, and the following three indicators were recorded: data upload delay, model synchronization time, and prediction response jitter. The test platform was configured as an edge server and a cloud platform.

Table 3 shows the edge deployment performance of

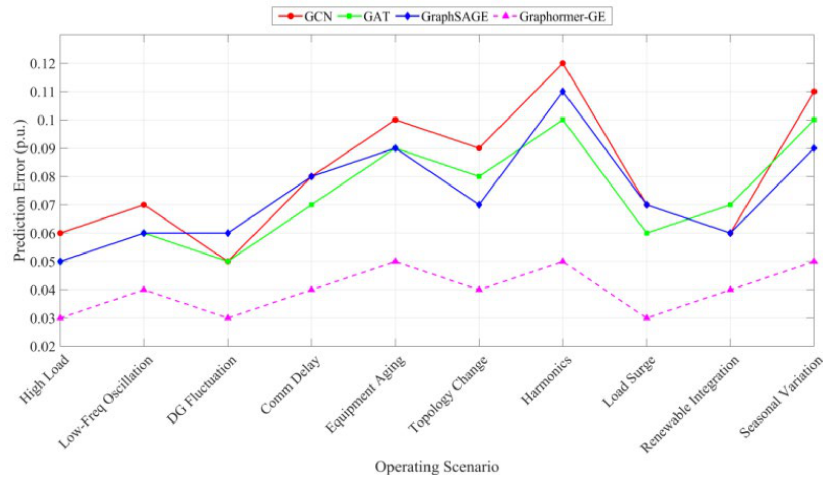


FIGURE 7. Multi-scenario generalization capability evaluation

TABLE 3. Statistics of model response delays under different network environments

Network Environment	Data Upload Delay (ms)	Model Synchronization Time (ms)	Prediction Response Jitter (ms)
Local Area Network	35	40	5
5G Network	80	95	12
Public Internet	220	250	40

Graphormer-GE in three typical network environments: LAN, 5G network, and public Internet. Specific indicators include data upload delay, model synchronization time, and prediction response jitter. In the LAN environment, the three indicators are only 35 ms, 40 ms, and 5 ms, respectively, showing extremely high real-time performance and stability. In the 5G network, due to the uncertainty brought by wireless transmission, the data upload delay rises to 80 ms. The model synchronization time increased to 95 ms, and the prediction response jitter also increased to 12 ms. When deployed in a public Internet environment with greater bandwidth fluctuations and higher latency, the three types of latency increased significantly, reaching 220 ms, 250 ms, and 40 ms, respectively. Despite this, Graphormer-GE was able to maintain acceptable inference stability, without significant fluctuations in prediction results or service interruptions.

From the data trend, the proposed method has better adaptability than traditional deployment methods in network-constrained environments, thanks to the network-grid collaborative dynamic graph update mechanism it introduced. This mechanism can perceive the communication link status in real time and dynamically adjust the graph attention weight distribution accordingly, so that when the data transmission between edge nodes is unstable, it can still maintain a relatively consistent modeling output. In the public Internet environment, although the data upload delay is as high as 220 ms, due to the adaptive optimization of the graph structure, the prediction response jitter does not increase exponentially, but is controlled within 40 ms, indicating that the model has certain fault tolerance and robustness. This design makes Graphormer-GE not only

suitable for high-quality LAN deployment, but also adaptable to wide-area edge computing scenarios, meeting the distributed operation and coordinated control requirements of smart grids in multi-network environments.

MODEL RESPONSE DELAY SENSITIVITY

To further quantify the response stability of the model in network-constrained scenarios, the sensitivity of the model output results to communication delays in different network environments was tested. The experiment was conducted in a local area network, 5G network, and public Internet environment, and the trend of the model prediction error changing with data upload delay was recorded.

Figure 8 is the model response delay sensitivity, which shows the trend of power flow prediction error change of four methods, GCN, GAT, GraphSAGE and Graphormer-GE, under different communication delay conditions. The horizontal axis represents the average data upload delay, and the vertical axis represents the average absolute error of the power flow prediction, which is used to measure the prediction stability of the model under the influence of delay. As the delay increases, the errors of GCN, GAT, and GraphSAGE all show a clear upward trend. The error of GCN has reached 0.069 p.u. at 250 ms, GAT is 0.0548 p.u., and GraphSAGE is 0.0615 p.u., while Graphormer-GE shows significantly better robustness, with an error of only 0.0348 p.u. at the same delay, with the smallest increase. This shows that Graphormer-GE has stronger adaptability in network-constrained scenarios.

Graphormer-GE introduces a dynamic graph update strategy based on the quality of the communication link, which can adjust the attention weight distribution according to the

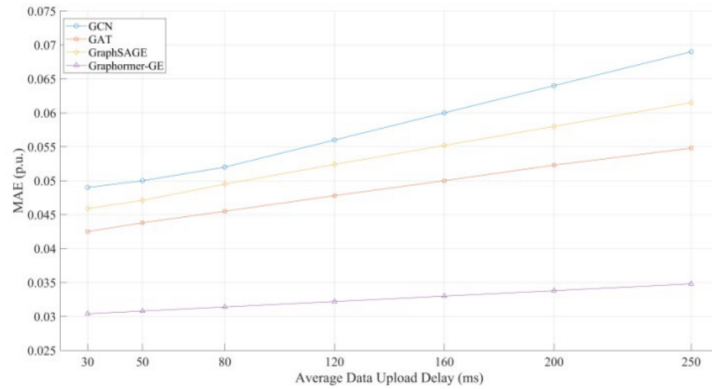


FIGURE 8. Model response delay sensitivity

real-time delay information and suppress the accumulation of information deviation caused by data upload delay. The topology-adaptive geometric encoding mechanism integrates the electrical distance information between grid nodes, so that the model can maintain stable feature expression capabilities when the graph structure changes or the communication quality decreases. Compared with other methods, it is difficult to cope with the input uncertainty caused by communication fluctuations, resulting in a significant increase in error with delay.

IV. CONCLUSION

By constructing a high-dimensional dynamic graph structure that integrates electrical characteristics and communication status, this paper combines the topological adaptive geometric coding mechanism, hybrid spatiotemporal attention modeling, and a network-grid collaborative dynamic graph update strategy to achieve high-precision and rapid response prediction of the grid operation status, a capability that directly addresses the needs of power-intensive sectors like the industry by ensuring the stable voltage and frequency essential for the continuous and precise operation of manufacturing machinery. Experimental results show that this method exhibits superior prediction accuracy and robustness in IEEE 39, 118, and 300 node systems. In the IEEE 300 node system, it has significant advantages in prediction accuracy (MAE is 0.037), response speed (Post-fault Error is 0.07), and edge deployment adaptability in the public Internet environment, including data upload delay (220 ms), model synchronization time (250 ms), and prediction response jitter (40 ms). In future work, this method can explore its distributed deployment application in larger-scale actual power grids, particularly focusing on industrial feeder lines connected to major manufacturing hubs. Furthermore, integrating the joint modeling capabilities of multi-energy coupling systems will be crucial to promoting its engineering implementation in the fields of intelligent scheduling, resilience improvement, and edge intelligence, thereby optimizing energy efficiency and ensuring power quality control specifically for the high-demand, flexible

load profiles characteristic of the modern industry.

AUTHOR CONTRIBUTIONS

Conceptualization – Huiqin Li and Xingbo Zhang; methodology – Huiqin Li and Xingbo Zhang; investigation – Huiqin Li and Xingbo Zhang; writing-original draft preparation – Huiqin Li and Xingbo Zhang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] M. Fan, Z. Li, T. Ding, L. Huang, F. Dong, Z. Ren, et al., "Uncertainty evaluation algorithm in power system dynamic analysis with correlated renewable energy sources," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5602-5611, 2021, doi: 10.1109/TPWRS.2021.3075181.
- [2] Y. Liu, A. K. Singh, J. Zhao, A. S. Meliopoulos, B. Pal, M. A. bin Mohd Ariff, et al., "Dynamic state estimation for power system control and protection," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5909-5921, 2021, doi: 10.1109/TPWRS.2021.3079395.
- [3] Y. Li, C. Yu, M. Shahidehpour, T. Yang, Z. Zeng, and T. Chai, "Deep reinforcement learning for smart grid operations: Algorithms, applications, and prospects," *Proceedings of the IEEE*, vol. 111, no. 9, pp. 1055-1096, 2023, doi: 10.1109/JPROC.2023.3303358.
- [4] T. Wang, S. Wang, S. Ma, J. Guo, and X. Zhou, "An extended continuation power flow method for static voltage stability assessment of renewable power generation-penetrated power systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 71, no. 2, pp. 892-896, 2022, doi: 10.1109/TCSII.2022.3209335.
- [5] T. Blenk and C. Weindl, "Fundamentals of state-space based load flow calculation of modern energy systems," *Energies*, vol. 16, no. 13, pp. 4872, 2023, doi: 10.3390/en16134872.
- [6] B. Li, J. Liu, Y. Su, Q. Zhang, D. Zhang, and Z. Liu, "A Dynamic Power Flow Calculation Method for the "Renewable Energy-Power Grid-Transportation Network" Coupling System," *Energies*, vol. 18, no. 7, pp. 1567, 2025, doi: 10.3390/en18071567.
- [7] W. Liao, B. Bak-Jensen, J. R. Pillai, Y. Wang, and Y. Wang, "A review of graph neural networks and their applications in power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 345-360, 2021, doi: 10.35833/MPCE.2021.000058.

- [8] X. Liu, Y. Zhang, Z. Zhen, F. Xu, F. Wang, and Z. Mi, "Spatio-temporal graph neural network and pattern prediction based ultra-short-term power forecasting of wind farm cluster," *IEEE Transactions on Industry Applications*, vol. 60, no. 1, pp. 1794-1803, 2023, doi: 10.1109/TIA.2023.3321267.
- [9] D. Beinert, C. Holzhüter, J. M. Thomas, and S. Vogt, "Power flow forecasts at transmission grid nodes using graph neural networks," *Energy and AI*, vol. 14, Art. no. 100262, 2023, doi: 10.1016/j.egyai.2023.100262.
- [10] R. Ma, H. Sun, and L. Qin, "An emergency coordinated control strategy to improve the transient stability of a single-ended distribution network with flexible interconnection channel blocking," *Sensors*, vol. 23, no. 20, pp. 8467, 2023, doi: 10.3390/s23208467.
- [11] M. Y. Mehmood, A. Oad, M. Abrar, H. M. Munir, S. F. Hasan, H. A. U. Muqet, et al., "Edge computing for IoT-enabled smart grid," *Security and communication networks*, vol. 2021, no. 1, Art. no. 5524025, 2021, doi: 10.1155/2021/5524025.
- [12] A. Althoubi, R. Alshahrani, and H. Peyravi, "Delay analysis in IoT sensor networks," *Sensors*, vol. 21, no. 11, pp. 3876, 2021, doi: 10.3390/s21113876.
- [13] X. Feng, Y. Zhang, A. Gao, and Q. Hu, "Research on AUV Multi-Node Networking Communication Based on Underwater Electric Field CSMA/CA Channel," *Biomimetics*, vol. 9, no. 11, pp. 653, 2024, doi: 10.3390/biomimetics9110653.
- [14] M. J. E. Alam, K. M. Muttaqi, and D. Sutanto, "Battery energy storage to mitigate rapid voltage/power fluctuations in power grids due to fast variations of solar/wind outputs," *Ieee Access*, vol. 9, pp. 12191-12202, 2021, doi: 10.1109/ACCESS.2021.3051283.
- [15] J. B. Hansen, S. N. Anfinsen, and F. M. Bianchi, "Power flow balancing with decentralized graph neural networks," *IEEE Transactions on Power Systems*, vol. 38, no. 3, pp. 2423-2433, 2022, doi: 10.1109/TPWRS.2022.3195301.
- [16] D. Chen, M. Chen, P. Wu, M. Wu, T. Zhang, and C. Li, "Two-stream spatio-temporal GCN-transformer networks for skeleton-based action recognition," *Scientific Reports*, vol. 15, no. 1, pp. 4982, 2025, doi: 10.1038/s41598-025-87752-8.
- [17] X. Zhan, L. Kou, M. Xue, J. Zhang, and L. Zhou, "Reliable long-term energy load trend prediction model for smart grid using hierarchical decomposition self-attention network," *IEEE Transactions on Reliability*, vol. 72, no. 2, pp. 609-621, 2022, doi: 10.1109/TR.2022.3174093.
- [18] Z. Tao, C. Wu, Y. Liang, K. Wang, and L. He, "LW-GCN: A lightweight FPGA-based graph convolutional network accelerator," *ACM Transactions on Reconfigurable Technology and Systems*, vol. 16, no. 1, pp. 1-19, 2022, doi: 10.1145/3550075.
- [19] J. Han, N. Liu, and J. P. S. Catalão, "Optimization of distribution network and mobile network with interactive balance of flexibility and power," *IEEE Transactions on Power Systems*, vol. 38, no. 3, pp. 2512-2524, 2022, doi: 10.1109/TPWRS.2022.3180792.
- [20] J. Wang and J. Ren, "Graphormer based contrastive learning for recommendation," *Applied Soft Computing*, vol. 159, Art. no. 111626, 2024, doi: 10.1016/j.asoc.2024.111626.
- [21] S. Yu, Y. Wang, L. Chen, X. Zhang, and J. Li, "3D hand pose and mesh estimation via a generic Topology-aware Transformer model," *Frontiers in Neurorobotics*, vol. 18, Art. no. 1395652, 2024, doi: 10.3389/fnbot.2024.1395652.
- [22] Z. Wang, Y. Hu, R. Zhong, B. Hu, D. Cheng, Y. Feng, et al., "St-Graphormer: spatio-temporal graph transformer for end-to-end traffic forecasting," *The Journal of Supercomputing*, vol. 81, no. 8, pp. 1-25, 2025, doi: 10.1007/s11227-025-07354-z.
- [23] C. Wang, Y. Wang, Z. Ding, T. Zheng, J. Hu, and K. Zhang, "A transformer-based method of multienergy load forecasting in integrated energy system," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2703-2714, 2022, doi: 10.1109/TSG.2022.3166600.
- [24] H. Toroslu I., "The Floyd-Warshall all-pairs shortest paths algorithm for disconnected and very sparse graphs," *Software: Practice and Experience*, vol. 53, no. 6, pp. 1287-1303, 2023, doi: 10.1002/spe.3188.
- [25] N. A. M. Nordin, S. S. R. Shariff, S. S. Supadi, and I. Masudin, "Modelling the Shortest Path for Inner Warehouse Travelling Using the Floyd-Warshall Algorithm," *Mathematics*, vol. 12, no. 17, pp. 2698, 2024, doi: 10.3390/math12172698.
- [26] B. Dokhniak and V. Khavalko, "Graph neural networks for traffic flow prediction: innovative approaches, practical usage, and superiority in spatio-temporal forecasting," *Bulletin of National Technical University "KhPI". Series: System Analysis, Control and Information Technologies*, pp. (2 (14)):34-39, 2025, doi: 10.20998/2079-0023.2025.02.05.
- [27] G. Chen, S. Chen, D. Li, and C. Chen, "A hybrid deep learning air pollution prediction approach based on neighborhood selection and spatio-temporal attention," *Scientific Reports*, vol. 15, no. 1, pp. 3685, 2025, doi: 10.1038/s41598-025-88086-1.
- [28] Y. Gao, L. Liang, T. Su, and M. Pan, "An embedded spatiotemporal hybrid model integrating multi-graphs and attention-driven fusion for single-and multi-site photovoltaic power forecasting," *Energy Conversion and Management*, vol. 336, Art. no. 119897, 2025, doi: 10.1016/j.enconman.2025.119897.
- [29] Z. Cheng, C. Yan, F. X. Wu, and J. Wang, "Drug-target interaction prediction using multi-head self-attention and graph attention network," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 4, pp. 2208-2218, 2021, doi: 10.1109/TCBB.2021.3077905.
- [30] B. Wang and J. Wang, "ST-MGAT: Spatio-temporal multi-head graph attention network for Traffic prediction," *Physica A: Statistical Mechanics and its Applications*, vol. 603, Art. no. 127762, 2022, doi: 10.1016/j.physa.2022.127762.
- [31] C. Shao, K. Li, T. Qian, M. Shahidepour, and X. Wang, "Generalized user equilibrium for coordination of coupled power-transportation network," *IEEE Transactions on Smart Grid*, vol. 14, no. 3, pp. 2140-2151, 2022, doi: 10.1109/TSG.2022.3206511.
- [32] L. N. Liu and G. H. Yang, "Distributed optimal economic environmental dispatch for microgrids over time-varying directed communication graph," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1913-1924, 2021, doi: 10.1109/TNSE.2021.3076526.
- [33] S. Chen, Y. Zhang, and Q. Yang, "Multi-task learning in natural language processing: An overview," *ACM Computing Surveys*, vol. 56, no. 12, pp. 1-32, 2024, doi: 10.1145/3663363.
- [34] A. Wang, Y. Ye, X. Song, S. Zhang, and J. J. Yu, "Traffic prediction with missing data: A multi-task learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4189-4202, 2023, doi: 10.1109/TITS.2022.3233890.
- [35] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE transactions on knowledge and data engineering*, vol. 34, no. 12, pp. 5586-5609, 2021, doi: 10.1109/TKDE.2021.3070203.