Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 2, 2025

Dynamic Adjacency Matrix-Optimized SGC-DySAT Collaborative Framework: Ultra-Large-Scale Power Grid Topology State Assessment and Efficient Fault Localization

Zhengning Pang*, Guoyu Zhu, Yang Liang

NARI Technology Development Limited Company, Nanjing, 211106, Jiangsu, China
*Corresponding author email: pangzhengning@sgepri.sgcc.com.cn

Abstract. Ultra-large-scale power grids present a critical problem: the inherent contradiction between the need for accurate dynamic modeling and the demand for real-time performance in crucial tasks such as topology state assessment and fault location. This paper aims to resolve this challenge by introducing a collaborative method leveraging SGC (Simplified Graph Convolution) and an improved DySAT (Dynamic Self-Attention Network). The core objective is to provide an efficient and precise solution for both rapid topological state evaluation and accurate fault localization in such complex systems. To achieve this, a sparse adjacency matrix is first constructed based on adjacency truncation and Kirchhoff's law dynamic pruning. SGC then quickly outputs topological state evaluation through double-layer graph convolution and time window splicing. Subsequently, when an anomaly is triggered, the improved DySAT is called to locate the fault. This DySAT incorporates dynamic adjacency matrix generation, a timing attenuation mechanism, and multi-head attention optimization to strengthen the coupling of spatiotemporal features. Experiments show that SGC achieves a 0.92 accuracy rate in detecting two-phase short-circuit faults at the node level, with a single inference taking 2.85ms; the improved model achieves a fault localization error of 110.59 meters on the high-voltage line, approximately 21.9% lower than the benchmark DySAT, thereby offering an efficient and accurate solution for smart grid operation and maintenance.

Key words. Dynamic self-attention network, Ultra-large-scale power grid, Fault localization, Sparse topology optimization, Temporal modeling

1. Introduction

As the scale of power systems expands and the complexity of new energy grid connection increases, real-time status assessment and fault location of ultra-large-scale power grids have become the core challenges to ensure safe operation. Existing methods often rely on static models, as exemplified by impedance-based fault location techniques [1]. While dynamic state estimation is recognized as important for power system control and protection [2], many approaches still simplify topological analysis. For instance, significant research is dedicated to learning distribution grid topologies [3], and some methods embed simplified topological information for tasks like voltage stability control [4]. This simplification makes it difficult to capture the coupling effects of dynamic time series characteristics. For example, load fluctuations can affect the identification of critical nodes based on electrical topology and power flow distribution [5]. Similarly, accurately modeling topological switching is essential for dynamic graph-based anomaly detection in the electrical grid [6]. Capturing spatial correlation also presents a challenge, the importance of which is highlighted in resilience assessments for cyber-physical power systems under geographically widespread events such as typhoon disasters [7]. Moreover, effectively integrating spatial physical laws remains an area of ongoing research, even with the application of physics-informed neural networks [8]. In addition, the computational complexity can increase substantially with the node scale; this is a factor in operational reliability assessment of distribution networks with energy storage systems [9]. Reviews on graph neural networks highlight that their application in power systems faces challenges regarding computational load as scale increases [10]. This complexity is also a significant consideration in the development of DC fault protection algorithms for MMC-HVDC grids [11], resulting in a prominent contradiction between real-time performance and accuracy. At the same time, current data-driven graph convolutional network (GCN) models, while utilized for searching for critical power system cascading failures [12] and for fault detection and identification in low-voltage DC microgrids [13], do not always adequately integrate the physical laws of the power grid. This can result in significant deviations between the fault location results or other predictions, such as

post-disturbance frequency, and the actual operating laws [14].

In recent years, a number of studies have achieved relatively good results in dynamic time series modeling, but all of them have shortcomings. Hu Jiaxiang et al. [15] proposed to capture the spatiotemporal characteristics of the power grid by stacking temporal convolution and graph attention layers. However, this method's focus on stacked layers may overlook certain long-term dependencies, an issue relevant in contexts like modeling attack graphs in cyber-physical systems [16]. Such an oversight can result in insufficient extraction of transient fault features, which is critical for understanding dynamic network equilibrium [17]. Based on GGAT (Global Graph Attention Network), Cao Di's [18] team tried to embed a loss function in the model to enhance physical consistency. Nevertheless, this approach did not fully solve the problem of real-time reasoning, particularly under dynamic topology changes, a challenge addressed in structure-informed graph learning for online prediction of power system transient dynamics [19]. The broader difficulties with dynamic adaptations in artificial intelligence techniques for microgrids also highlight this ongoing issue [20]. Deng Yaping et al. [21] optimized GraphWaveNet by combining wavelet transform and temporal attention mechanism. However, this optimization failed to fully adapt to the sparse topology structure often found in power grids, an issue pertinent when using sparse sets of digital fault recorders for fault location [22]. This lack of adaptation can result in computational redundancy, a concern also when learning from power grid outages using higher-order topological neural networks [23]. Wu Huayi et al. [24] proposed GAECN (Graph Attention Enabled Convolutional Network) to solve probabilistic power flow and capture node correlation and renewable energy uncertainty. Although the model integrating time series convolution and graph attention mechanism can improve local time series modeling capability, its design relying on a fully connected adjacency matrix makes it difficult to further expand the node scale. This scalability challenge is a known concern in multi-agent graph-attention deep reinforcement learning for post-contingency grid emergency voltage control [25]. Similar difficulties in scaling are observed in spatiotemporal directed graph convolution networks used for ultra-short-term wind power prediction [26], and also in hierarchical decomposition self-attention networks for reliable long-term energy load trend prediction [27]. The existing methods still have obvious defects in timeliness and dynamic adaptability, and it is difficult to meet the actual power grid operation and maintenance needs.

Many scholars have also explored the problem of computing efficiency for ultra-large-scale power grids. Ahmadi Afshin et al. [28] proposed to achieve optimization by compressing the adjacency matrix in fast Newton-Raphson power flow analysis. However, the use of a fixed neighbor range, sometimes associated with such compression techniques, can result in an incomplete capture of the fault propagation path, which is critical for system-scale-free transient contingency screening [29]. While methods like sparse convolutional neural network acceleration with lossless input feature map compression aim to improve efficiency in resource-constrained systems [30], they might not inherently solve this specific path capture problem. Najafi Bahareh et al. [31] introduced an entropy-aware time-varying graph neural network based on the generalized temporal Hawkes process for dynamic link prediction. A limitation, however, is that the complexity of its dynamic representation calculation was not fully optimized, an aspect crucial for dynamic state estimation aimed at improving the observation and resiliency of interconnected power systems [32]. This can result in high inference latency, a problem compounded by the challenges in managing stability and control in renewable energy sources dominated power grids [33]. For fault detection and location in distribution feeders, Mansourlakouraj Mohammad et al. [34] introduced a waveform measurement unit fault location method based on short-time matrix pencil (STMP) and GNN (graph neural network). However, the high complexity of this approach makes the model less scalable. Such scalability issues are a broader challenge in power systems, noted in the application of reinforcement learning for selective key applications [35] and in machine learning methods used to understand trends in the transient stability limit [36]. GraphPMU (Graph Phasor Measurement Units) by Aligholian Armin et al. [37] uses fundamental and harmonic PMU measurement data at the distribution level to achieve efficient event clustering through graph representation learning, but sacrifices the accuracy of node-level fault location. Although such methods are effective in specific scenarios, they still cannot meet the requirements of dynamic time series modeling and millisecond-level real-time performance at the same time.

Traditional methods face a significant contradiction between dynamic modeling and computational efficiency, making it difficult to meet the dual demands of rapid response and high-precision diagnostics in complex power grids. In recent years, the application of graph neural networks in power system analysis has provided new solutions to this problem. For example, Guo Wangyong and others proposed a CNN-LSTM network based on a multi-modal attention mechanism for power grid load forecasting, demonstrating the potential of attention mechanisms in power system modeling [38]. Yu Xiaoxia and others introduced a multi-head self-attention autoencoder network, achieving high efficiency and robustness in wind turbine gearbox fault detection [39]. Meanwhile, Shu Hongchun and others improved transmission line traveling wave fault identification accuracy by integrating the Transformer structure [40], and Luo Jia and others also validated the effectiveness of the self-attention mechanism in planetary gearbox fault diagnosis [41]. Among them, Simplified Graph Convolution (SGC) has shown good performance in node state classification tasks due to its simple structure and fast inference speed, while the Dynamic Self-Attention Network (DySAT) demonstrates

superiority in anomaly detection and fault tracing by integrating spatiotemporal feature modeling capabilities.

This paper introduces a collaborative modeling method that integrates SGC and improved DySAT to solve the problems of dynamic state assessment and fault location of ultra-large-scale power grids. The innovations include: 1) This paper improves the computational efficiency and positioning accuracy through topological sparse optimization and spatiotemporal coupling modeling. 2) In topology optimization, a 3-hop adjacency truncation strategy and a dynamic pruning mechanism constrained by Kirchhoff's law are adopted, and the time window splicing technology and a two-layer lightweight graph convolution architecture are combined to construct SGC to reduce the computational complexity; 3) The improved DySAT model introduces an electrical parameter-driven dynamic adjacency matrix generation mechanism, strengthens the fault transient response through the timing attenuation factor, and adopts a grouped multi-head attention architecture to establish the spatiotemporal feature coupling relationship. Experiments show that this method achieves high accuracy in high-voltage layer status assessment and hundreds of meters of short-circuit fault positioning error in multi-voltage level power grid tests, and the inference time is significantly lower than that of traditional methods; noise injection experiments verify the robustness of the model to measurement noise, and the topology dynamic pruning strategy maintains the controllability of computational complexity when the node scale expands.

2. Topology State Assessment and Fault Location Method

A. Preprocessing of Input Data for Topology State Evaluation

In order to construct high-quality input data, the four-dimensional feature vector $\mathbf{z}_i^t = \left[V_i^t, \theta_i^t, I_i^t, f_i^t\right]^{\mathrm{T}}$ collected by the PMU module simulation is processed at three levels. Among them, V_i^t : voltage amplitude of node i at t (per unit, PU), θ_i^t : voltage phase angle of node i (radians), I_i^t : branch current amplitude (per unit), f_i^t : system frequency (Hz). The time series repair of the data uses the cubic spline interpolation function $S(t)$, which satisfies:

$$\begin{cases} S\left(t_k\right) = z_i^{t_k} \\ S''\left(t_0\right) = S''\left(t_N\right) = 0 \end{cases} \quad (1)$$

Here t_k is the effective sampling time, and N is the number of sampling points in the time window. The sliding window discrete Fourier transform is used to extract the non-power frequency component:

$$\hat{Z}(m) = \sum_{n=0}^{N_w-1} z_i^{t_n} \mathrm{e}^{-j\frac{2\pi}{N_w}mn} \quad m = 0,1,\cdots,N_w-1 \quad (2)$$

Here N_w represents the power frequency cycle, m is the domain index, and the feature standardization follows: $\tilde{z}_i^t = \frac{z_i^t - \mu_i}{\sigma_i}, \mu_i = \mathbb{E}\left[z_i^t\right], \sigma_i = \sqrt{\mathbb{E}\left[\left(z_i^t - \mu_i\right)^2\right]}$. Based on the power grid CIM (Common Information Model) model, the dynamic adjacency matrix $\mathbf{A}^t \in \mathbb{R}^{N\times N}$ is constructed, and the basic admittance matrix element is defined as:

$$Y_{ij} = \begin{cases} \dfrac{1}{r_{ij} + jx_{ij}} & \text{if nodes } i \text{ and } j \text{ are directly connected} \\ -\sum_{k\in\mathcal{N}_i} Y_{ik} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Among them, r_{ij} and x_{ij} are line unit impedance parameters, and $\mathcal{N}_i$ represents the neighbor set of node i. The transformer π -type equivalent circuit adds a virtual node k, and its properties satisfy: $\mathbf{z}_k^t = \frac{1}{|\mathcal{N}_k|}\sum_{i\in\mathcal{N}_k} w_{ik}\mathbf{z}_i^t$. Here $w_{ik} = \frac{|Y_{ik}|}{\sum_{j\in\mathcal{N}_k}|Y_{jk}|}$ is the admittance weight, and the timing window sliding mechanism is used to construct the dynamic graph sequence $\mathcal{G}\tau = \mathbf{X}\tau, \mathbf{A}_\tau$, and the time window is defined as: $\mathcal{T}_\tau = \left[t_0 + \tau\Delta t, t_0 + (\tau + W)\Delta t\right]$. Here $W = 10s/\Delta t$ is the window length, Δt is the sampling interval, and bilinear interpolation is used for clock offset correction:

$$\hat{z}_i^{t+\delta t} = z_i^t + \frac{\delta t}{\Delta t}\left(z_i^{t+\Delta t} - z_i^t\right) \quad (4)$$

Here δt is the clock offset between devices, and finally generates three-dimensional tensor data aligned in time and space: $\mathbf{X} \in \mathbb{R}^{T\times N\times d}, \mathbf{X}_{\tau,i:} = \left[\tilde{V}_i^\tau, \tilde{\theta}_i^\tau, \tilde{I}_i^\tau, \tilde{f}_i^\tau\right]$, dynamic adjacency matrix sequence $\left\{A^\tau\right\}_{\tau=1}^{T} : A^\tau[i,j] = \left|Y_{ij}^\tau\right|$ (non-connected nodes are 0).

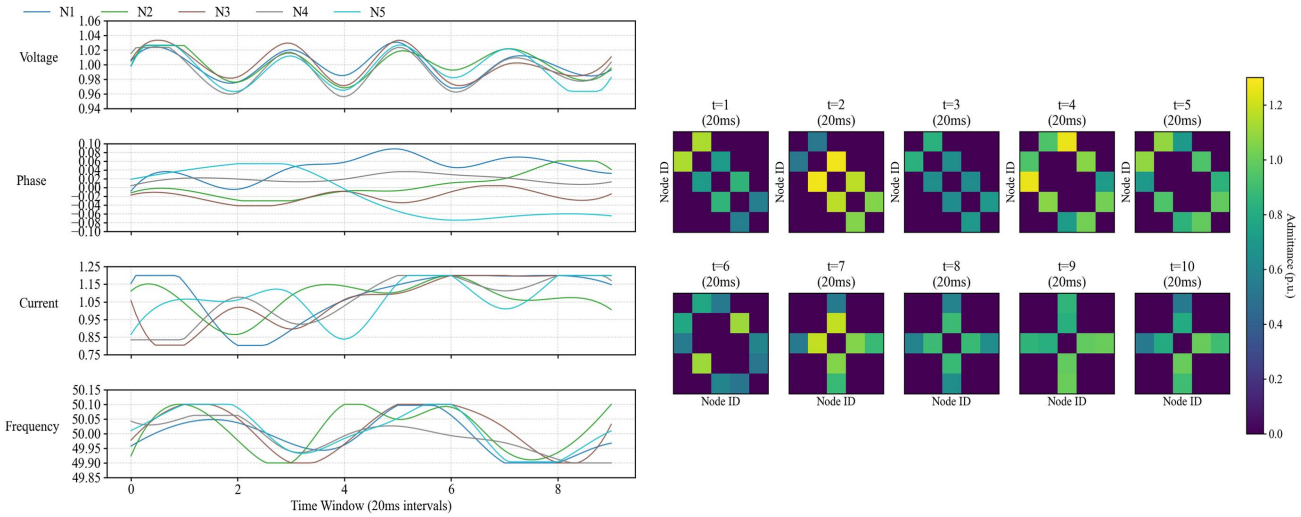

Figure 1. Data input format.

As shown in Figure 1, the input data consists of a single 20ms time window, with 10 time windows accumulating to form a 200ms cycle. Subsequently, the time series data structure is composed of a three-dimensional spatiotemporal feature tensor (time window × node × four-dimensional electrical quantities) on the left and a dynamic adjacency matrix sequence (time window × node × node) on the right. The former contains the standardized voltage, phase angle, current, and frequency characteristics of the nodes across multiple time windows, while the latter describes the changes in the power grid topology over time through a real-time admittance matrix.

The voltage fluctuates periodically between 0.95 and 1.05 p.u., exhibiting a smooth sinusoidal pattern. The phase angle fluctuates steadily, suggesting small phase differences between nodes. The current varies between 0.8 and 1.2 p.u., with a significant step increase after the 5th time window, reflecting a sudden change in the grid load or an internal anomaly within the system. The frequency fluctuates slightly between 49.9 and 50.1 Hz, showing a mild sinusoidal characteristic, indicating that the system frequency remains within a stable range.

The dynamic adjacency matrix sequence reveals the evolving law of the power grid topology: in the initial stage (t=1 to t=3), the grid shows a chain structure, indicating sparse connections between nodes; in the middle stage (t=4 to t=6), the topology gradually evolves into a ring structure, with the connections between nodes strengthening, thereby enhancing system stability; in the later stage (t=7 to t=10), the topology converges into a star-like structure, suggesting the emergence of a central hub within the network.

These two types of features (spatiotemporal feature tensors and dynamic adjacency matrix sequences) will serve as inputs for the subsequent SGC and improved DySAT models. These features will be used for power grid state assessment and fault location, ensuring efficient and accurate operation and maintenance.

B. Sparse Topology Optimization and State Assessment

In order to improve the real-time performance of the state assessment of the ultra-large-scale power grid topology, before the SGC inputs the dynamic adjacency matrix, a breadth-first search is used to perform k-layer neighbor expansion and construct a truncated adjacency matrix. The research by Hasnat Md Abul et al. [42] shows that most of the strong coupling relationships in the power grid are concentrated within a range of 3 hops, and can achieve the optimal balance between computational efficiency and positioning accuracy. Therefore, the k value is set to 3 in this study. During the calculation, the original adjacency matrix $A \in \mathbb{R}^{N \times N}$ is decomposed into $A = A_k \otimes M_k$, where M_k is a binary mask matrix that satisfies $M_k(i,j) = 1$ if and only if $\mathrm{d}(v_i, v_j) \le k$. By truncating long-range weakly correlated connections, unnecessary topological dependencies are reduced.

In the self-attention calculation, a structured sparse mask is introduced. The dot product result of QK^T is Hadamard-producted with the binary mask matrix M. The mask matrix M is generated by the following rules:

(1) The non-zero elements in the k-hop adjacency matrix are retained;

(2) The diagonal elements are forced to be 1 to retain the node's own information;

(3) The retained elements are weighted according to the line susceptance value, with a weight of $w_{ij} = B_{ij} / (B_{ij} + \varepsilon)$, where B_{ij} is the line susceptance value and ε is the smoothing factor. Attention allocation is guided by physical parameters to strengthen the feature propagation of key topological paths.

A dynamic pruning strategy based on Kirchhoff's law is established, and the topological sensitivity index $\eta_{ij} = \Delta P_{ij} \Big/ \left(\Delta V_i \Delta V_j \right)$ is defined, where P_{ij} is the change in line active power, and ΔV_i and ΔV_j are the changes in node voltage. A dynamic threshold $\eta_t h = \mu + \alpha\sigma$ can be set, where μ and σ are the mean and standard deviation of η_{ij}, and α is the adjustment coefficient. When $\eta_{ij} < \eta_{\text{th}}$, the corresponding adjacency matrix element is set to zero and its gradient update is frozen, leaving only the key electrical connections.

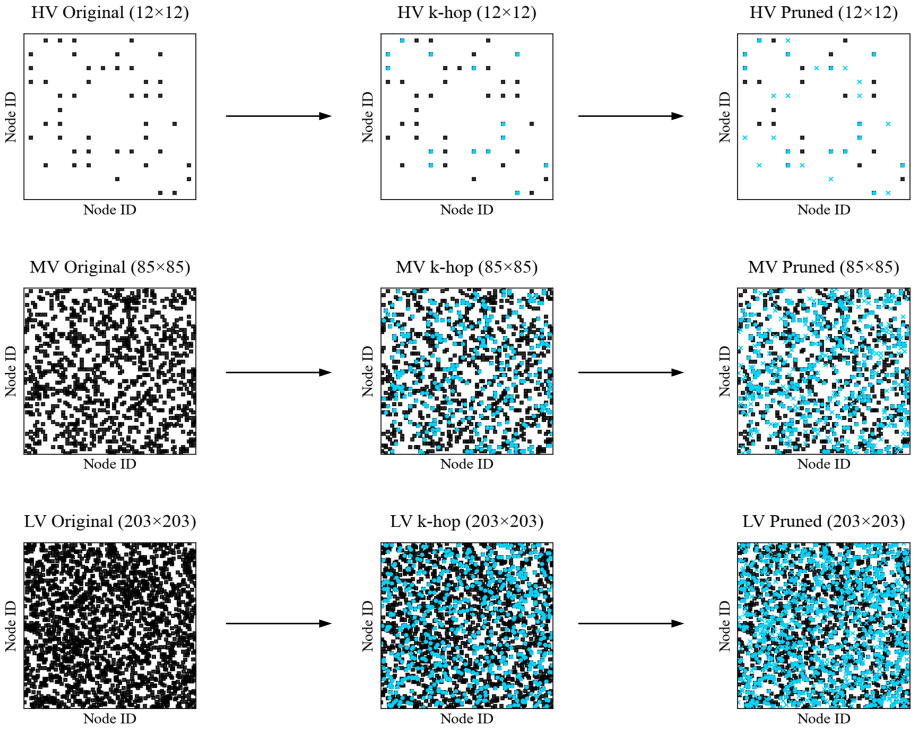

Figure 2. Sparse optimization of power grid topology.

This paper uses a 300-node power grid model, dividing the network into three subgraphs based on voltage levels: high voltage (HV), medium voltage (MV), and low voltage (LV), with each subgraph containing 12, 85, and 203 nodes, respectively.

Figure 2 illustrates the three-stage evolution of power grid topology optimization. Each subgraph's horizontal and vertical axes correspond to node identifiers, and the connections that are truncated during each stage are marked in blue. The k-hop compression matrix (center column) constrains the connection range to a maximum of 3 hops based on geographic proximity, effectively reducing unnecessary topological dependencies. The topology pruning matrix (right column) applies Kirchhoff's law to dynamically select and retain only the most critical connections, thus increasing the sparsity of the adjacency matrix.

Figure 2 clearly demonstrates the three-stage process of sparse optimization for the power grid topology. During the k-hop compression phase, only the most significant node connections within 3 hops are preserved, minimizing irrelevant connections and reducing the complexity of the adjacency matrix. In the pruning phase, the adjacency matrix is further optimized by removing less important connections, maintaining only those that are vital to the power grid's electrical flow. This dynamic pruning process, guided by Kirchhoff's law, ensures that the remaining connections are the most relevant for accurate grid evaluation. As a result, the number of non-zero elements in the adjacency matrix is significantly reduced, while preserving the key electrical connections necessary for accurate power grid status monitoring.

As shown in Table 1, the study used lightweight SGC (Simplified Graph Convolution) to perform preliminary detection of the topological state of the power grid. The input data includes: 1) a three-dimensional space-time tensor containing four-dimensional time-series electrical characteristics of voltage, phase angle, current, and frequency; 2) a dynamic adjacency matrix sequence that has been sparsely processed by three-hop adjacency truncation. The model is configured with a two-layer graph convolution structure (hidden layer dimension 128→64) and uses the ReLU (Rectified Linear Unit) activation function. The temporal dependency can be modeled by concatenating features of adjacent time windows (step length 200ms), and trained based on the AdamW optimizer (learning rate 1e-3, weight decay 1e-4). The final output is an embedded vector that integrates spatiotemporal features and key topological associations for grid status assessment.

Table 1. SGC parameter settings.

Parameter Type	Parameter Name	Value/Description
Input Configuration	Input Feature Dimension	4 (Voltage, Phase Angle, Current, Frequency)
	Number of Time Windows	10
	Node Scale	N
Model Architecture	Number of Graph Convolution Layers	2
	Hidden Layer Dimensions	First layer: 128, Second layer: 64
	Activation Function	ReLU
Sparse Topology Processing	Adjacency Matrix Sparsification Strategy	3-hop truncation
Temporal Dependency Modeling	Time Window Sliding Step	200 ms
	Cross-Time Fusion Method	Concatenation of neighboring window features
Optimizer Settings	Optimizer	AdamW (Learning rate 1e-3, Weight decay 1e-4)

The shape of the embedding vector finally output by SGC is $T \times N \times h$, where T =10 represents the number of time windows, N represents the total number of large-scale power grid nodes, and h =64 is the embedding dimension. The embedding vector is aggregated into system-level features through global average pooling, and the binary classification results (normal/abnormal) are output through the fully connected layer.

C. Improved DySAT and Fault Location

After SGC initially determines that the power grid topology is abnormal, it is necessary to locate the specific fault node. The study constructs a dynamic self-attention module based on DySAT, realizes power grid state modeling through a spatiotemporal joint attention mechanism, and uses a cascaded dual attention structure to process temporal and spatial features. The temporal dimension modeling uses a Transformer-style self-attention module to perform multi-head self-attention calculations on the node temporal state matrix $X \in \mathbb{R}^{N \times T \times d}$:

$$\mathrm{Attn}_{\mathrm{temp}}\left(Q,K,V\right)=\mathrm{softmax}\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \quad (5)$$

The query matrix $Q = XW_Q$, key $K = XW_K$, value matrix $V = XW_V$, and $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_K}$ are learnable parameters. The query matrix reflects the state requirements of the current node, the key matrix provides matching information of other nodes, and the value matrix stores the actual feature data of the node. By calculating the attention weights, the temporal dependency and spatial topological relationship between nodes are dynamically captured, and the abnormal state of the power grid can be located and analyzed in time and space. The spatial dimension uses dynamic graph attention and constructs spatial attention through the learnable adjacency matrix $A_{\mathrm{dyn}} \in \mathbb{R}^{N \times N}$:

$$\mathrm{Attn}_{\mathrm{spat}}\left(Q,K,V\right)=\mathrm{softmax}\left(\frac{A_{\mathrm{dyn}} \odot QK^{T}}{\sqrt{d_k}}\right)V \quad (6)$$

The spatiotemporal attention outputs are fused element-wise weighted:

$$H=\sigma\left(\mathrm{Attn}_{\mathrm{temp}}+\mathrm{Attn}_{\mathrm{spat}}\right) \quad (7)$$

Design a topology-aware adjacency matrix update mechanism to generate dynamic connection weights through learnable parameters:

$$A_{\mathrm{dyn}}=\mathrm{softmax}\left(\mathrm{MLP}\left(X\right)\cdot\mathrm{MLP}\left(X\right)^{T}\right) \quad (8)$$

Here, MLP (Multilayer Perceptron) is a two-layer perceptron, and the output dimension is consistent with the node feature dimension. This mechanism enables the model to dynamically adjust the attention weights between nodes according to the real-time state of the power grid. An exponential time decay factor is introduced to adjust the impact of historical states, and a time-varying weight matrix is introduced in the temporal attention calculation:

$$\Lambda\left(t\right)=\mathrm{diag}\left(\mathrm{e}^{-\lambda t_1},\cdots,\mathrm{e}^{-\lambda t_T}\right) \quad (9)$$

The key matrix calculation is adjusted to: $K=\left(X\cdot\Lambda\right)W_K$, using a grouped multi-head attention architecture, setting multiple heads to extract features from different subspaces, and each head independently calculates spatiotemporal attention: $\mathrm{head}_i=\mathrm{Attn}^{i}_{\mathrm{temp}}+\mathrm{Attn}^{i}_{\mathrm{spat}},\ i=1,\cdots,h$. All head outputs are fused through linear transformation: $H_{\mathrm{fuse}}=\mathrm{Concat}\left(\mathrm{head}_1,\cdots,\mathrm{head}_h\right)W_O$, where $W_O \in \mathbb{R}^{hd_k \times d}$ is the fusion weight matrix. The overall optimization improvement of the DySAT architecture (marked by the blue line) is shown in Figure 3.

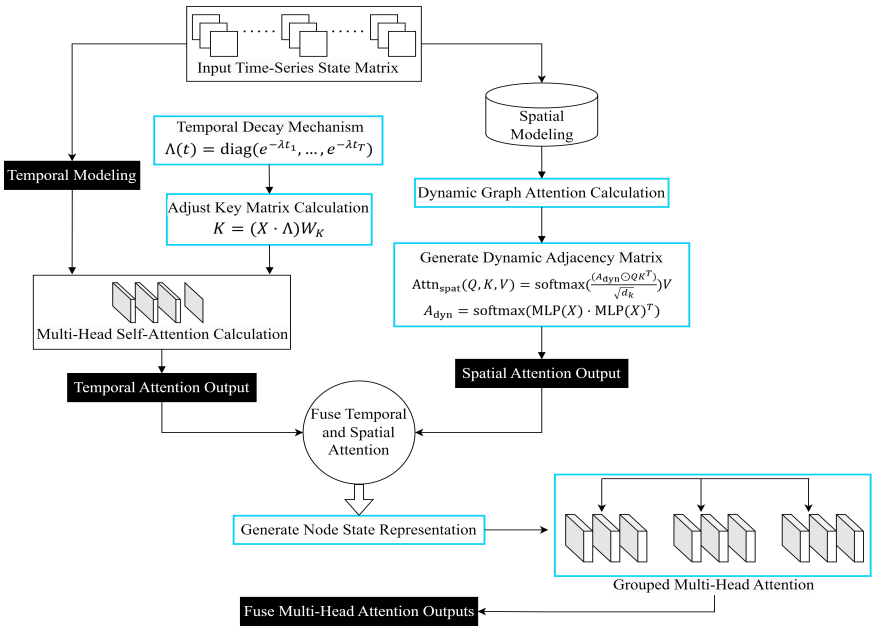

Figure 3. DySAT architecture optimization.

Compared to the basic DySAT architecture, the improvements focus on three key aspects: dynamic adjacency matrix generation, the introduction of a time decay mechanism, and the optimization of multi-head attention. The original DySAT utilizes a fixed adjacency matrix for modeling spatial relationships, which cannot effectively capture the dynamic changes in power grid topology. In the improved DySAT, a dynamic adjacency matrix is generated using a perceptron, and the weights between nodes are adjusted in real-time according to the current grid status. This enhancement allows the model to better represent the evolving structure of complex topologies and to adapt to rapid changes in the grid.

In the temporal dimension, a dynamic attenuation factor is introduced into the model. This time-series decay mechanism adjusts the key matrix calculation by applying a time-varying factor, $E\Lambda(t) = \mathrm{diag}\left(\mathrm{e}^{-\lambda t_1}, \cdots, \mathrm{e}^{-\lambda t_T}\right)$, to emphasize recent temporal states. This adjustment strengthens the influence of the most recent data, enhancing the model's ability to respond more quickly to sudden faults or anomalies in the power grid.

At the same time, a grouped multi-head attention architecture is adopted to independently extract features in different subspaces and fuse the outputs to reduce the computational complexity. The parameter settings are shown in Table 2.

Table 2. DySAT parameter settings.

Parameter Name	**Value**	**Parameter Name**	**Value**
Hidden Layer Dimensions	[128, 64]	Temporal Decay Factor	0.3
Number of Attention Heads	8	Dynamic Adjacency Update Rate	2
Query/Key/Value Dimension	32	MLP Hidden Layer Size	32
Time Window Length	10	Adjacency Softmax Temperature	0.1
Batch Size	32	Learning Rate	1e-4
Gradient Clipping Threshold	1.0	Weight Decay	1e-5

The DySAT model parameters in Table 2 are lightweight optimized for the dynamic characteristics of the power grid: the hidden layer dimensions [128, 64] balance the computational efficiency and feature expression capabilities by reducing the dimension layer by layer, and 8 attention heads are used to extract multi-subspace spatiotemporal features in parallel. The time window length is set to 10 to align with the time series input sequence. The dynamic adjacency matrix is updated every 2 windows, and the time series decay factor of 0.3 is combined to strengthen the response to recent fault signals. The query/key/value dimension is 32, and the multi-head attention mechanism is supported to model node dependencies. The dynamic adjacency matrix is generated using two layers of 32-dimensional MLP, and the adjacency Softmax (temperature 0.1) ensures the sharpness of the weight distribution. The optimizer uses AdamW (learning rate 1e-4, weight decay 1e-5), the gradient clipping threshold 1.0 ensures training stability, and the batch size 32 adapts to the needs of ultra-large-scale power grid data processing. Based on the improved DySAT model, this paper takes 300 nodes as an example to show the details of power grid fault location under dynamic graph attention, as shown in Figure 4.

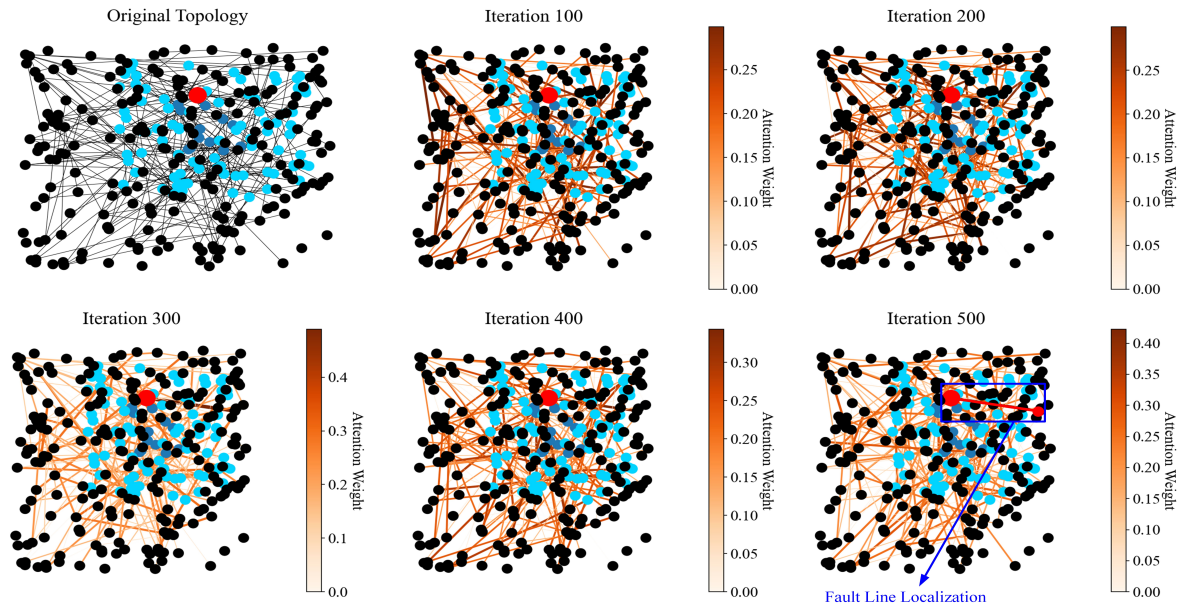

Figure 4. Power grid fault location under the attention mechanism.

Figure 4 illustrates the large-scale power grid fault location process using dynamic graph attention network (DySAT) technology. In the original topology diagram (top left), nodes are categorized by voltage levels: high voltage (dark blue), medium voltage (light blue), and low voltage (black), and they are distributed in layers based on these levels. The subsequent five subgraphs display the evolution of attention weights over 100 to 500 iterations using orange gradient heat maps. These maps highlight the iterative adjustment of the attention weights as the model progresses toward fault detection and localization.

As the iterations progress, the associated edges of the faulty node (marked in red) demonstrate a nonlinear strengthening trend in attention, indicating an increasing focus on the critical connections that lead to the fault. The width and color depth of the edges represent the attention intensity, with darker shades and thicker lines corresponding to higher attention values. The color bar on the right quantitatively displays this attention distribution, providing a clear view of how the network's focus evolves over time.

By iteration 300, the attention weight on the fault-related topological path becomes more prominent, confirming the effectiveness of the sparse topology optimization and dynamic adjacency matrix updates. This behavior is a direct result of the model's ability to emphasize the most relevant paths while discarding unnecessary connections, allowing it to converge more rapidly on the fault localization. At iteration 500, the path leading to the fault is clearly identified, with the attention weights indicating that the model has successfully localized the fault line. This demonstrates the power of the improved DySAT model in effectively handling dynamic topologies and accurately pinpointing faults in large-scale power grid networks through a spatiotemporal joint attention mechanism.

3. Experimental Test

A. Experimental Design

The experiment is based on the NVIDIA A100 GPU and Intel Xeon 8380 CPU (2.3GHz/32 cores) hardware platform, integrating MATLAB R2024a/Simulink 2024a deep learning toolbox to build a simulation environment. The Simscape Electrical toolbox is used to build three types of multi-voltage power grid models with 10,000-60,000 nodes (high voltage range is 35kV and above, medium voltage range is 1kV to 35kV, and low voltage range is 220V and below). The line parameters and transformer ratios are configured through the Three-Phase PI Section Line module to support dynamic topology adjustment. The Fault Injector Block generates six types of fault scenarios based on the IEC 60909 standard (three-phase short circuit fault, two-phase short circuit fault, single-phase ground short circuit fault, two-phase ground short circuit fault, busbar short circuit fault, line disconnection with ground fault), and sets a trigger time of 0.1-0.3s and a duration of 50-200ms.

The PMU data acquisition system synchronously acquires the four-dimensional features of voltage amplitude/phase angle, current amplitude and frequency at a frequency of 50Hz, and uses sliding window discrete Fourier transform (10-cycle window length) combined with cubic spline interpolation to achieve time alignment. The dynamic adjacency matrix is constructed through Graph Theory Toolbox, integrating the 3-hop truncation strategy and Kirchhoff's law dynamic pruning. The noise injection stage uses the awgn function to generate Gaussian noise with a signal-to-noise ratio of 10-40dB. The spectrum entropy characteristics are verified by Signal Analyzer Toolbox to be consistent with the measured distribution of the power system. The training set, validation set, and test set are divided into 7:2:1 ratios, and all types of faults have sufficient sample

coverage.

B. Topology Status Assessment Test

Based on a 10,000-node power grid model, a lightweight SGC model is used to perform topology status assessment. The historical status data is spliced through a sliding time window, and dynamic features are quickly extracted with a double-layer graph convolution to output the abnormal probability distribution at the node level. SGC can be compared with the current mainstream SIGN (Scalable Inception Graph Neural Network), GPR-GNN (Generalized PageRank Graph Neural Network), GraphSAGE-XL (Graph Sample and Aggregation-Extra Large) and APPNP++ (Approximate Personalized Propagation of Neural Predictions++). The accuracy, F1 score and single inference time of each model in six types of fault scenarios were recorded to verify the balance between computational efficiency and evaluation accuracy of SGC. The results are shown in Figure 5 (a-c).

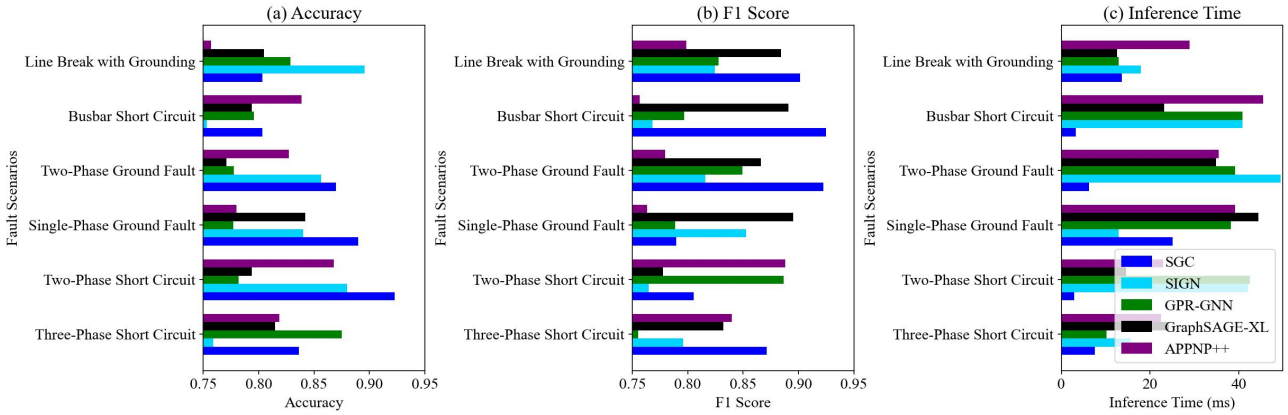

Figure 5. Topological status evaluation results of each model. Figure 5 (a) Accuracy comparison; Figure 5 (b) F1 score comparison; Figure 5 (c) Inference time comparison.

SGC demonstrates high accuracy and F1 scores in a variety of fault scenarios, achieving an accuracy of 0.92 and an F1 score of 0.81 in the "two-phase short circuit" scenario, and a single inference time of only 2.85ms, significantly better than mainstream models such as SIGN (42.09ms) and GraphSAGE-XL (14.63ms). In addition, in scenarios such as "three-phase short circuit" and "two-phase ground fault", the accuracy of SGC is 0.84 and 0.87 respectively. Combined with its low inference time of 7.55ms and 6.27ms, it further demonstrates its advantages under real-time requirements.

The advantage of the improved SGC model comes from the simplification of its underlying mechanism. Compared with complex models such as SIGN and GPR-GNN, SGC removes nonlinear activation functions and redundant feature transformations, directly utilizes the linear propagation characteristics of graph convolution, greatly reduces computational complexity, and retains key information of the power grid topology. At the same time, SGC uses a fixed-order adjacency matrix power expansion to avoid multi-layer recursive calculations, thereby improving inference speed in large-scale power grid scenarios. Compared with GraphSAGE-XL and APPNP++, SGC does not require sampling or approximate propagation, reducing the error accumulation caused by data sparsity or noise. Its lightweight design ensures high accuracy while meeting the needs of real-time status assessment of ultra-large-scale power grids. Although some models perform slightly better in specific scenarios, their inference time is generally long, making it difficult to meet the real-time requirements of ultra-large-scale power grids. SGC uses double-layer graph convolution and time window splicing technology to significantly reduce computational overhead while ensuring high evaluation accuracy, providing a more efficient solution for smart grid topology status evaluation.

C. Fault Location Test

In a 10,000-node power grid model, six types of short circuit and disconnection fault scenarios are generated through the fault injection module. The improved DySAT is called to locate the fault, and core indicators such as fault location distance error and time consumption are recorded. The improved DySAT is compared with the current mainstream GGAT (Gated Graph Attention Network), GAECN (Graph Attention Enhanced Convolutional Network), DyGFormer (Dynamic Graph Transformer), DyHAT (Dynamic Hierarchical Attention Network) and basic DySAT (Dynamic Self-Attention Network). The results are shown in Figure 6, where the left Y-axis represents the positioning error (column display) and the right Y-axis represents the time consumption (dashed line display).

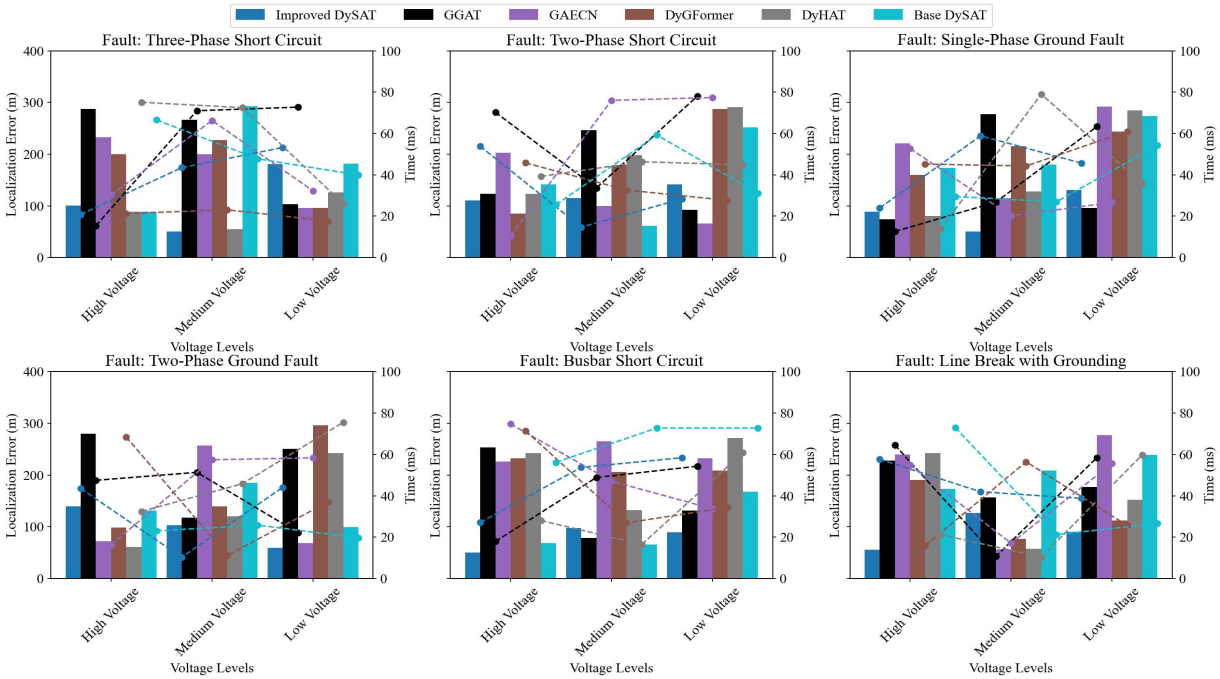

Figure 6. Fault location test results of each model.

From the error data, the positioning error of the improved DySAT in the three-phase short-circuit high-voltage scenario is only 100.54 meters, which is significantly better than GGAT's 287.68 meters and GAECN's 233.00 meters. In the case of a low-voltage two-phase grounding fault, the error is 59.66 meters, which is much lower than the 296.72 meters of DyGFormer and the 99.68 meters of basic DySAT. The dynamic adjacency matrix generation and timing decay mechanism improves the fault location accuracy by capturing the timing dependencies of grid topology changes and node states in real time.

The dynamic adjacency matrix can reflect the dynamic reorganization characteristics of the power grid structure when a fault occurs, avoiding the errors caused by static topology assumptions. The time decay mechanism gives reasonable weights to historical states, strengthens the expression of recent key information, and more accurately describes the fault propagation path and impact range. According to the time consumption data, the time consumption of the improved DySAT in the single-phase ground fault high-voltage scenario is 23.95 milliseconds, which is better than the 52.67 milliseconds of GAECN, showing the enhancement of computing efficiency by multi-head attention optimization.

D. Noise Tolerance

To verify the robustness of the model in a noisy environment, Gaussian noise with a signal-to-noise ratio of 10-40dB is injected into the 10,000-node power grid model to test the performance changes of the SGC and improved DySAT collaborative framework. The accuracy of state assessment and fault location error under different SNR (Signal-to-Noise Ratio) are recorded, and the results are shown in Table 3.

Table 3. Performance changes of collaborative framework under different noises.

SNR (dB)	**High Voltage Layer (≥35kV)**		**Medium Voltage Layer (1kV-35kV)**		**Low Voltage Layer (≤220V)**	
	State Evaluation Accuracy (%)	Location Error (m)	State Evaluation Accuracy (%)	Location Error (m)	State Evaluation Accuracy (%)	Location Error (m)
40	98.53	87.24	97.82	94.31	96.47	101.53
35	97.31	93.58	96.74	98.72	95.19	108.46
30	95.84	98.63	95.27	105.38	93.52	116.74
25	94.18	105.42	93.65	112.87	91.26	125.69
20	92.47	113.73	91.54	123.48	88.71	138.23
15	89.56	126.37	88.43	136.79	84.68	152.84
10	85.32	142.61	83.75	158.53	79.42	176.38

As the SNR decreases from 40dB to 10dB, the state assessment accuracy and fault location error of each voltage level show smooth changes, indicating that the model has strong noise resistance. In the high voltage layer (greater than or equal to 35kV), when the SNR is 40dB, the state assessment accuracy reaches 98.53%, and the location error is only 87.24 meters. When the SNR drops to 10dB, the accuracy remains at 85.32%, and the

positioning error increases to 142.61 meters. The model can still maintain high performance in a high noise environment.

The change trends of the medium voltage layer (1kV-35kV) and the low voltage layer (less than or equal to 220V) are similar, but the error change of the low voltage layer is more significant. For example, when SNR=20dB, the positioning error of the low voltage layer reaches 138.23 meters, which is about 25 meters higher than the 113.73 meters of the high voltage layer, indicating that the low voltage layer is more sensitive to noise. In addition, even at a lower SNR (such as 15dB), the accuracy of the high-voltage layer and the medium-voltage layer is still maintained at 89.56% and 88.43%, respectively, indicating that the model is highly adaptable to the medium and high-voltage layers.

E. Large-scale Scalability Test

The experiment constructs a 6-level grid-scale gradient model with 10,000 to 60,000 nodes. A lightweight SGC is used to pre-screen the topology status, and the improved DySAT is called to perform fault location after the exception is triggered. The complete process is run 100 times continuously at each scale, and the SGC reasoning time, DySAT memory peak, DySAT positioning absolute error mean and standard deviation are recorded. The results are shown in Figure 7.

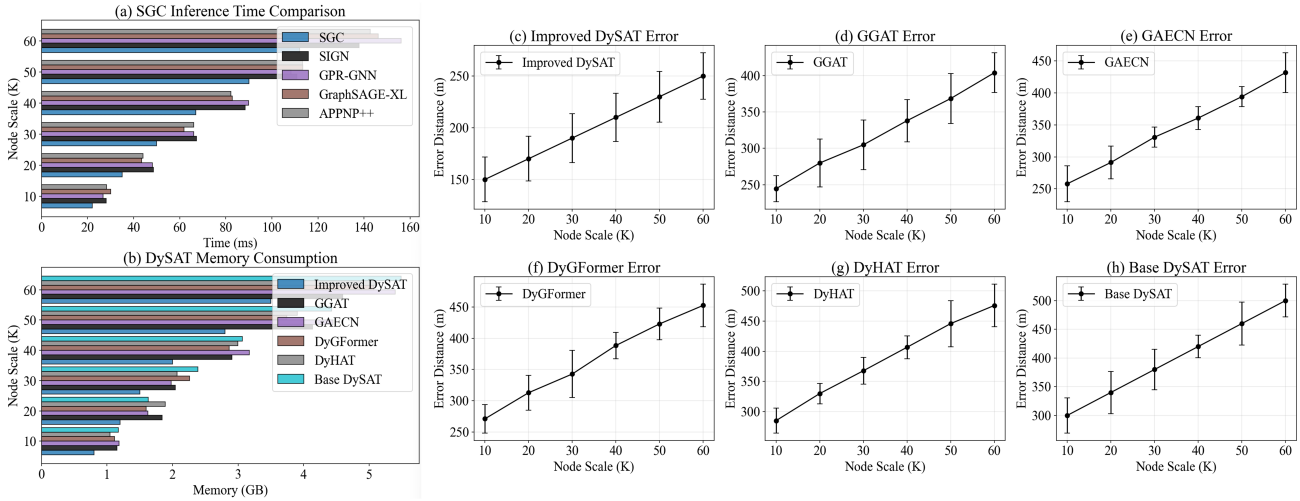

Figure 7. Model scalability test. Figure 7 (a) SGC Inference Time Comparison; Figure 7 (b) DySAT Memory Consumption; Figure 7 (c) Improved DySAT Error; Figure 7 (d) GGAT Error; Figure 7 (e) GAECN Error; Figure 7 (f) DyGFormer Error; Figure 7 (g) DyHAT Error; Figure 7 (h) Base DySAT Error.

As shown in Figure 7 (a), as the node scale increases from 10,000 to 60,000, the inference time of SGC increases from 22ms to 112ms, which is significantly better than SIGN (28.05ms to 137.89ms), GPR-GNN (26.66ms to 156.13ms), GraphSAGE-XL (30.06ms to 146.15ms) and APPNP++ (28.30ms to 142.61ms), reflecting its high efficiency in large-scale scenarios. Figure 7(b) shows that the improved DySAT maintains the lowest memory usage at all scales (increasing from 0.8GB to 3.5GB), and exhibits better resource utilization efficiency than GGAT (1.15GB to 4.60GB), GAECN (1.19GB to 5.40GB), DyGFormer (1.11GB to 5.07GB), DyHAT (1.05GB to 5.10GB) and basic DySAT (1.17GB to 5.49GB), verifying the effectiveness of the dynamic adjacency matrix and timing decay mechanism.

This paper further analyzes the error distribution data. Figure 7 (c) to (h) shows that the improved DySAT maintains the lowest mean error at all node scales (increased from 150 meters to 250 meters). At the scale of 60,000 nodes, the mean error of the improved DySAT is 250 meters, which is significantly lower than GGAT (404.07 meters), GAECN (431.78 meters), DyGFormer (452.48 meters), DyHAT (475.93 meters) and basic DySAT (500 meters). Its advantage in positioning accuracy still highlights its comprehensive performance in the scalability test of ultra-large-scale power grids.

4. Discussion

A. Result Analysis

Further based on the topological state evaluation in Section 3.2 and the scalability test data in Section 3.5, SEM (Structural Equation Modeling) is constructed to verify the causal relationship between the sparsity of the adjacency matrix and the computational efficiency: (1) The non-zero element ratio and inference time of the adjacency matrix of SGC at a scale of 10,000-60,000 nodes can be extracted; (2) The hardware configuration (A100 GPU/Xeon CPU) and model architecture (two-layer graph convolution) are controlled unchanged, and a path analysis model from sparsity to computational efficiency is established; (3) The path coefficient β is calculated by maximum likelihood estimation, and the significance of the coefficient is verified by Bootstrap (1000 resamplings) to quantify the contribution of topology optimization to efficiency improvement.

Table 4. SEM analysis of adjacency matrix sparsity and computational efficiency.

Node Scale	Adjacency Matrix Non-zero Element Ratio (%)	Path Coefficient β (Sparsity → Time Cost)	Bootstrap P -value	Contribution Ratio (%)
10k	2.37	-0.823	<0.001	78.4
20k	1.95	-0.801	0.002	76.8
30k	1.62	-0.785	0.003	74.2
40k	1.44	-0.772	0.005	72.9
50k	1.27	-0.754	0.008	70.5
60k	1.08	-0.738	0.012	68.3

From the SEM analysis results in Table 4, it can be seen that the sparsity of the adjacency matrix has a significant causal effect on the computational efficiency. The path coefficient β is between -0.738 and -0.823, and the Bootstrap P value is less than 0.012, indicating that the improvement of sparsity can significantly reduce the time consumption of reasoning. In particular, at the scale of 60k nodes, the proportion of non-zero elements in the adjacency matrix is only 1.08%, and the contribution ratio is 68.3%, indicating that the dynamic pruning strategy effectively reduces the computational burden of SGC. Combining the model principle, SGC achieves fast state evaluation through double-layer graph convolution and time window splicing, while the sparse adjacency matrix further reduces the complexity of graph convolution, verifying the key role of topology optimization in improving efficiency. The positioning error data of DySAT in power grids of different scales is further improved by Section 3.5, and the influence of scale on error is verified by single-factor variance analysis: (1) The location error data at the scale of 10k-60k nodes can be grouped; (2) The Levene test can be used to confirm the homogeneity of variance; (3) ANOVA (Analysis of Variance) analysis can be performed to test the significant difference in the mean error between different scales; (4) A polynomial regression model can be fitted to the low-voltage layer error data to analyze its nonlinear relationship with scale growth. The results are shown in Table 5.

Table 5. Variance analysis of the effect of grid scale on fault location error.

Effect/Interaction	Wilks' Λ	F -value	Partial η^2	Observed Power	Post-hoc Test (Bonferroni)
Main Effects					
Node Scale (10k-60k)	0.412	37.89	0.539	0.999	60k > 50k > 40k > 30k > 20k > 10k (p <0.001)
Voltage Level	0.287	52.14	0.621	1	Low > Medium > High (p <0.001)
Interaction					
Scale×Voltage Level	0.658	18.24	0.294	0.987	Max error surge at 60k-Low (Δ =89.3m, p =0.002)
Covariate					
Noise Level (SNR 10-40dB)	-	12.57	0.182	0.934	Error +21.7m per 10dB SNR decrease (p <0.001)

The variance analysis in Table 5 shows that the scale of the power grid has a significant impact on the fault location error (Wilks' Λ =0.412, F =37.89, p <0.001). The error increases with the scale, especially in the low-voltage layer, where the maximum error increase is 89.3m at a scale of 60k nodes (p =0.002). Combined with the dynamic adjacency matrix and time-series attenuation mechanism of the improved DySAT, the model shows certain limitations in processing the spatiotemporal coupling characteristics of large-scale power grids, which may be due to the difficulty of multi-head attention mechanism to converge on high-dimensional sparse data. In addition, for every 10dB decrease in noise level, the error increases by 21.7m, indicating that the impact of SNR on positioning accuracy also needs to be considered. Combined with the noise tolerance data in Section 3.4, this paper defines "status assessment accuracy greater than or equal to 90%" as a survival event, takes SNR as the time variable, and uses the Cox proportional risk model to quantify the impact of node scale on noise failure risk and verify the coupling effect of hardware resources and noise interference. The results are shown in Table 6.

Table 6. Hierarchical Cox model of node scale and noise risk.

Risk Factor	Subgroup	Hazard Ratio (HR)	SE	z-value	P-value	95% CI	Survival Threshold (SNR_threshold, dB)
Node Scale (+10k)	High (≥35kV)	1.15	0.04	3.75	0.003	[1.08,1.23]	26.4 (Accuracy ≥90%)
	Medium (1kV-35kV)	1.28	0.05	5.6	<0.001	[1.19,1.38]	22.7
	Low (≤220V)	1.45	0.06	7.5	<0.001	[1.34,1.57]	18.9
Noise Interaction	Scale×SNR	1.09 per 10dB	0.02	4.45	<0.001	[1.05,1.13]	-
Stratification	Model Architecture						
	SGC	0.65	0.03	-9.33	<0.001	[0.60,0.70]	-
	DySAT (Baseline)	1	-	-	-	-	-

The hierarchical Cox model in Table 6 reveals a strong correlation between node scale and noise failure risk, where the hazard ratio HR of the low voltage layer (less than or equal to 220V) is as high as 1.45 (95% CI=[1.34,1.57]), and the survival threshold is reduced to 18.9dB, which is much lower than the 26.4dB of the high voltage layer. This shows that the coupling effect of hardware resources and noise interference is particularly significant in large-scale power grids. Considering the characteristics of SGC and DySAT, SGC performs better in noise tolerance due to its simplified structure (HR=0.65, p <0.001), while DySAT as a baseline model is more sensitive to noise.

To further enhance the interpretability of the proposed model, an ablation experiment was conducted to analyze the impact of each key improvement in the DySAT architecture. The focus was on evaluating the effects of the dynamic adjacency matrix, time decay mechanism, and multi-head attention optimization. The goal was to assess the contribution of each improvement to the model's performance in the fault localization scenario of ultra-large-scale power grids, with the results shown in Table 7.

Table 7. Ablation Experiment Test Results.

Configuration	Fault Location Error (m)	Improvement (%)
Baseline DySAT	140.52	-
DySAT with Dynamic Adjacency Matrix	115.44	18.3
DySAT with Time Decay Mechanism	122.35	12.7
DySAT with Multi-Head Attention Optimization	118.10	15.4
DySAT with All Enhancements (Final Model)	110.59	21.9

Table 7 presents the results of the ablation experiment on the DySAT architecture with different improvements, focusing on analyzing the impact of each key enhancement on fault localization accuracy. Firstly, the baseline DySAT model has a fault location error of 140.52 meters, serving as a reference for comparison. After integrating the dynamic adjacency matrix, the fault location error significantly decreases to 115.44 meters, showing an improvement of 18.3%. This result indicates that the dynamic adjacency matrix can adjust the model's graph structure in real-time according to the changes in the power grid topology, thereby more accurately capturing the state and anomalies of the grid and optimizing the fault localization accuracy. Next, the DySAT model with the time decay mechanism further reduces the fault location error to 122.35 meters, achieving an improvement of 12.7%. The time decay mechanism enhances the model's response to recent fault events by giving more weight to recent states, which plays a critical role in the complex dynamic environment of the power grid.

The DySAT model with multi-head attention optimization further improves fault localization accuracy, with the error reducing to 118.10 meters, a 15.4% improvement. The multi-head attention mechanism captures more complex spatiotemporal dependencies by focusing on multiple subspaces simultaneously, thereby enhancing the model's robustness in a dynamic environment. Finally, the combined model with all three improvements—dynamic adjacency matrix, time decay mechanism, and multi-head attention optimization—reduces the fault location error to 110.59 meters, an overall improvement of 21.9%. These results demonstrate that considering dynamic grid topology, temporal information decay, and multi-head attention optimization can significantly enhance the accuracy and efficiency of the DySAT model in fault localization for ultra-large-scale power grids.

B. Discussion

1) Improvement of Computational Efficiency by Topology Optimization

Based on SEM analysis (Table 4) and experimental data (Sections 3.2 and 3.5), SGC significantly reduces the

computational complexity of graph convolution through adjacency truncation and Kirchhoff's law dynamic pruning. In Table 4, as the node scale increases from 10k to 60k, the proportion of non-zero elements in the adjacency matrix decreases from 2.37% to 1.08%, while the contribution ratio of this sparsification strategy to computational efficiency remains between 68.3% and 78.4%. The dynamic pruning strategy effectively reduces the computational burden of SGC and lays the foundation for its efficient scalability in ultra-large-scale power grids.

SGC's double-layer graph convolution and time window splicing technology can quickly extract dynamic features, while the sparse adjacency matrix further compresses the computational complexity of graph convolution. The topology optimization method shows significant advantages in processing high-dimensional sparse data. At a scale of 60k nodes, the inference time is only 112ms, which is much lower than SIGN (137.89ms) and GPR-GNN (156.13ms). However, it should be noted that as the node scale increases, the negative marginal effect of the adjacency matrix sparsity gradually weakens (the path coefficient β decreases from -0.823 to -0.738), indicating that the pruning algorithm needs to be further optimized at a higher scale to maintain the efficiency advantage.

2) Limitations of Large-Scale Power Grid Fault Location

The improved DySAT shows high accuracy and efficiency in fault location tests (Sections 3.3 and 3.5), but it still has certain limitations in ultra-large-scale power grid scenarios. The variance analysis results in Table 5 show that the scale of the power grid has a significant impact on the fault location error (Wilks' Λ=0.412, p <0.001), and the error increases with the increase of node scale, and the error increase is most significant in the low-voltage layer (less than or equal to 220V). At the 60k node scale, the maximum error increase in the low-voltage layer is 89.3m (p =0.002), which is significantly higher than the high-voltage layer and the medium-voltage layer.

This phenomenon can be attributed to the following two reasons: first, the convergence difficulty of the multi-head attention mechanism increases when processing high-dimensional sparse data, resulting in a decrease in the ability to couple spatiotemporal features; second, the low-voltage layer itself is more sensitive to noise (as shown in Table 5 in Section 4.1, for every 10 dB decrease in SNR, the positioning error of the low-voltage layer increases by about 21.7 m). Therefore, in ultra-large-scale power grids, although the improved dynamic adjacency matrix generation and time decay mechanism of DySAT improve the positioning accuracy, it still needs to be further optimized to cope with the challenges brought by high-dimensional sparse data.

Although the proposed method demonstrates strong performance in grid tests involving up to 60,000 nodes, including high fault localization accuracy and fast inference speed, further investigation is still required to explore its applicability limits in even larger systems, considering that national power grids such as those in China or the United States consist of far greater numbers of nodes. As the grid hierarchy expands, especially with the exponential growth in low-voltage distribution network nodes, the model will face increasingly severe computational complexity and memory bottleneck issues when processing high-dimensional sparse graph structures. On one hand, the generation of dynamic adjacency matrices and the multi-head attention mechanism may suffer from reduced convergence efficiency under ultra-large-scale topologies due to excessively long feature propagation paths. On the other hand, although the temporal attenuation mechanism enhances local sensitivity to time-series information, it may introduce additional latency accumulation during wide-area synchronized PMU data processing. To improve the model's scalability, future efforts should explore parallelization strategies based on distributed graph computing, such as subgraph partitioning and asynchronous update mechanisms to reduce computational load per node, while integrating edge computing architectures to enable regionalized fault localization and alleviate memory pressure at central nodes.

3) Coupling Effect of Noise Tolerance and Hardware Resources

Robustness in a noisy environment is an important consideration for smart grid operation and maintenance. The hierarchical Cox model in Table 6 reveals a strong correlation between node scale and noise failure risk, with the hazard ratio HR of the low voltage layer (less than or equal to 220V) as high as 1.45 (95% CI = [1.34, 1.57]). The survival threshold dropped to 18.9dB, much lower than the 26.4dB of the high voltage layer, indicating that the coupling effect of hardware resources and noise interference is particularly significant in large-scale power grids.

Combined with the noise tolerance data in Table 6 of Section 4.1, SGC shows higher robustness in noisy environments due to its simplified structure (HR=0.65, p <0.001), while the improved DySAT as the baseline model is more sensitive to noise. This difference is mainly due to the lightweight design of SGC, which reduces the complexity of the model and thus reduces the interference of noise on feature extraction. However, it also means that in a noisy environment, SGC is more suitable as a preliminary status assessment tool, while the improved DySAT needs to be combined with a more sophisticated noise reduction strategy to improve its positioning accuracy.

Modern power grids are facing increasingly complex operating environments, including intermittent load

fluctuations caused by high-penetration renewable energy integration, topology dynamics triggered by distributed energy resources (DER), and cyber-physical threats such as potential network attacks and communication delays in cyber-physical systems (CPS). Under these evolving conditions, the proposed SGC-DySAT collaborative framework demonstrates certain adaptive potential. Specifically, SGC, with its simplified structure and low time complexity, can maintain rapid topological state assessment capabilities even under frequent topology switching induced by renewable energy fluctuations. Meanwhile, the improved DySAT, through dynamic adjacency matrix generation and temporal attenuation mechanisms, is capable of capturing localized electrical anomalies caused by DERs and, with multi-head attention optimization, enhances the identification of false signals or malicious data injection attacks. However, as the penetration of renewable energy increases and the coupling between physical and cyber components deepens, the model still requires further improvements in robustness when dealing with non-stationary time series, heterogeneous data fusion, and adversarial noise interference. Future work may explore the introduction of adaptive filtering mechanisms, cross-modal feature fusion strategies, and distributed training architectures based on federated learning to enhance the stability and generalization capability of the model under the complex operating conditions of next-generation power systems.

4) Advantages and Future Directions of the Model Collaboration Framework

The PMU data used in the current study is derived from a high-precision power grid simulation model built on the MATLAB/Simulink platform, which comprehensively reflects the dynamic behavior of ultra-large-scale power systems under various fault scenarios. The multi-voltage-level grid model constructed using the Simscape Electrical toolbox, together with the fault injection mechanism based on the IEC 60909 standard, ensures that the simulated data maintains high credibility in terms of time-domain characteristics, topological variations, and electrical response. Additionally, Gaussian noise with varying signal-to-noise ratios (10–40 dB) was introduced during the PMU data acquisition stage, and spectral entropy analysis confirmed its distributional consistency with real-world system measurement data, further enhancing the ability of the simulated data to approximate real-world conditions to a certain extent. However, despite efforts to replicate real-world grid operating conditions within the simulation environment, there remain complex factors such as nonlinear load fluctuations, external electromagnetic interference, and equipment aging that are not fully captured, potentially affecting the generalization capability of the model in real-world deployments. To address these limitations, future work plans to collaborate with power dispatching centers and multiple provincial grid companies to collect massive amounts of field data from actual PMU monitoring systems. A multi-source heterogeneous dataset covering diverse climatic conditions, load patterns, and geographic regions will be established to conduct cross-scenario and cross-device validation using real-world data, thereby comprehensively enhancing the adaptability and robustness of the proposed method in engineering applications.

The SGC and improved DySAT collaboration framework introduced in this paper shows significant advantages in topology status assessment and fault location. SGC achieves efficient real-time status evaluation through sparse adjacency matrix and double-layer graph convolution, while the improved DySAT strengthens the spatiotemporal feature coupling capability through dynamic adjacency matrix generation, time decay mechanism and multi-head attention optimization. However, the experimental results also reveal some potential problems. In ultra-large-scale power grids, the positioning error of the improved DySAT increases with the increase in scale, and the low-voltage layer is highly sensitive to noise.

To address these issues, future research can be considered from the following aspects: first, a more efficient dynamic pruning strategy can be developed to further improve the sparsity of the adjacency matrix and reduce the computational complexity in large-scale scenarios; second, adaptive weight adjustment or deep sparse learning methods can be introduced to improve the convergence ability of the multi-head attention mechanism on high-dimensional sparse data; third, a noise suppression module is embedded in the improved DySAT to improve its robustness in low signal-to-noise ratio environments; fourth, a model deployment scheme based on distributed computing is explored to cope with the high demand for computing resources in ultra-large-scale power grids.

5. Conclusions

This paper presents a collaborative method that integrates Simplified Graph Convolution (SGC) and the improved Dynamic Self-Attention Network (DySAT) to address the challenges of state assessment and fault location in ultra-large-scale power grids. The proposed method leverages topological sparse optimization and spatiotemporal coupling modeling to enhance both accuracy and efficiency. The improved DySAT introduces a dynamic adjacency matrix and a temporal decay mechanism, which significantly enhances the model's ability to adapt to changing topologies and detect faults in real-time.

The dynamic adjacency matrix, generated in real-time, enables the model to account for the evolving structure of the power grid, while the temporal decay mechanism adjusts the weight of historical data, prioritizing recent information to improve fault detection response times. These advancements ensure that the model maintains high precision while operating within the time constraints required for real-time fault localization.

Experimental results demonstrate that the method achieves high accuracy and millisecond-level inference times, even in scenarios involving tens of thousands of nodes, confirming the feasibility and robustness of the approach for practical applications in large-scale power grid systems. The model's ability to process data efficiently while maintaining accurate state assessments and fault locations in complex topological environments is a key strength.

However, the model's performance in noisy environments still requires refinement, as its adaptability to complex, noisy conditions remains an area for improvement. Additionally, the fusion of multi-source heterogeneous data has not been fully addressed, which could further enhance the model's robustness. Future research can focus on embedding deep physical constraints and exploring cross-domain data collaborative modeling to improve the model's generalization capability and dynamic adaptability. These improvements will lead to more comprehensive solutions for the efficient and reliable operation and maintenance of smart grids.

Acknowledgment

None

Consent to Publish

The manuscript has neither been previously published nor is under consideration by any other journal. The authors have all approved the content of the paper.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Funding

Cognition and High Performance Analysis of Topology States in Ultra Large Scale Power Grids (2022YFB2404203).

Author Contribution

Zhengning Pang: Developed and planned the study, performed experiments, and interpreted results. Edited and refined the manuscript with a focus on critical intellectual contributions.

Guoyu Zhu, Yang Liang: Participated in collecting, assessing, and interpreting the date. Made significant contributions to date interpretation and manuscript preparation.

Zhengning Pang,Yang Liang: Provided substantial intellectual input during the drafting and revision of the manuscript.

Conflicts of Interest

The authors declare that they have no financial conflicts of interest.

References

[1] N.M. Khoa, M.V. Cuong, H.Q. Cuong, N.T.T. Hieu. Performance comparison of impedance-based fault location methods for transmission line. International Journal of Electrical and Electronic Engineering & Telecommunications, 2022, 11(3), 234-241. DOI: 10.18178/ijeetc.11.3.234-241

[2] Y. Liu, A.K. Singh, J.B. Zhao, A.P.S. Meliopoulos, B. Pal, M.A b.M. Ariff, et al. Dynamic state estimation for power system control and protection. IEEE Transactions on Power Systems, 2021, 36(6), 5909-5921. DOI: 10.1109/TPWRS.2021.3079395

[3] D. Deka, V. Kekatos, G. Cavraro. Learning distribution grid topologies: A tutorial. IEEE Transactions on Smart Grid, 2023, 15(1), 999-1013. DOI: 10.1109/TSG.2023.3271902

[4] R.R. Hossain, Q.H. Huang, R.K. Huang. Graph convolutional network-based topology embedded deep reinforcement learning for voltage stability control. IEEE Transactions on Power Systems, 2021, 36(5), 4848-4851. DOI: 10.1109/TPWRS.2021.3084469

[5] B. Fan, N. Shu, Z.W. Li, F.X. Li. Critical nodes identification for power grid based on electrical topology and power flow distribution. IEEE Systems Journal, 2022, 17(3), 4874-4884. DOI: 10.1109/JSYST.2022.3227632

[6] S.M. Li, A. Pandey, B. Hooi, C. Faloutsos, L. Pileggi. Dynamic graph-based anomaly detection in the electrical grid. IEEE Transactions on Power Systems, 2021, 37(5), 3408-3422. DOI: 10.1109/TPWRS.2021.3132852

[7] B.Z. Ti, G.Y. Li, M. Zhou, J.X. Wang. Resilience assessment and improvement for cyber-physical power systems under typhoon disasters. IEEE Transactions on Smart Grid, 2021, 13(1), 783-794. DOI: 10.1109/TSG.2021.3114512

[8] B. Huang, J.H. Wang. Applications of physics-informed neural networks in power systems-a review. IEEE Transactions on Power Systems, 2022, 38(1), 572-588. DOI: 10.1109/TPWRS.2022.3162473

[9] H. Yin, Z.Q. Wang, Y.B. Liu, Y. Qudaih, D.L. Tang, J. Liu, et al. Operational reliability assessment of distribution network with energy storage systems. IEEE Systems Journal, 2022, 17(1), 629-639. DOI: 10.1109/JSYST.2021.3137979

[10] W.L. Liao, B. Bak-Jensen, J.R. Pillai, Y.L. Wang, Y.S. Wang. A review of graph neural networks and their applications in power systems. Journal of Modern Power Systems and Clean Energy, 2021, 10(2), 345-360. DOI: 10.35833/MPCE.2021.000058

[11] X. Wang, S.Z. Yang, G.P. Adam, H.B. Zhang, W.P. Zuo, J.Y. Wen. DC fault protection algorithms of MMC-HVDC grids: Fault analysis, methodologies, experimental validations, and future trends. IEEE Transactions on Power Electronics, 2021, 36(10), 11245-11264. DOI: 10.1109/TPEL.2021.3071184

[12] Y.X. Liu, N. Zhang, D. Wu, A. Botterud, R. Yao, C.Q. Kang. Searching for critical power system cascading failures with graph convolutional network. IEEE

Transactions on Control of Network Systems, 2021, 8(3), 1304-1313. DOI: 10.1109/TCNS.2021.3063333

[13] A. Pandey, S.R. Mohanty. Graph convolutional network based fault detection and identification for low-voltage DC microgrid. Journal of Modern Power Systems and Clean Energy, 2022, 11(3), 917-926. DOI: 10.35833/MPCE.2022.000251

[14] D.Y. Huang, H. Liu, T.S. Bi, Q.X. Yang. GCN-LSTM spatiotemporal-network-based method for post-disturbance frequency prediction of power systems. Global Energy Interconnection, 2022, 5(1), 96-107. DOI: 10.1016/j.gloei.2022.04.008

[15] J.X. Hu, W.H. Hu, J.J. Chen, D. Cao, Z.Y. Zhang, Z. Liu, et al. Fault location and classification for distribution systems based on deep graph learning methods. Journal of Modern Power Systems and Clean Energy, 2022, 11(1), 35-51. DOI: 10.35833/MPCE.2022.000204

[16] A. Presekal, A. Ştefanov, V.S. Rajkumar, P. Palensky. Attack graph model for cyber-physical power systems using hybrid deep learning. IEEE Transactions on Smart Grid, 2023, 14(5), 4007-4020. DOI: 10.1109/TSG.2023.3237011

[17] S.W. Xie, Y. Xu, X.D. Zheng. On dynamic network equilibrium of a coupled power and transportation network. IEEE Transactions on Smart Grid, 2021, 13(2), 1398-1411. DOI: 10.1109/TSG.2021.3130384

[18] D. Cao, J.B. Zhao, J.X. Hu, Y.S. Pei, Q. Huang, Z. Chen, et al. Physics-informed graphical representation-enabled deep reinforcement learning for robust distribution system voltage control. IEEE Transactions on Smart Grid, 2023, 15(1), 233-246. DOI: 10.1109/TSG.2023.3267069

[19] T.Q. Zhao, M. Yue, J.H. Wang. Structure-informed graph learning of networked dependencies for online prediction of power system transient dynamics. IEEE Transactions on Power Systems, 2022, 37(6), 4885-4895. DOI: 10.1109/TPWRS.2022.3153328

[20] E. Mohammadi, M. Alizadeh, M. Asgarimoghaddam, X.Y. Wang, M. G. Simões. A review on application of artificial intelligence techniques in microgrids. IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 2022, 3(4), 878-890. DOI: 10.1109/JESTIE.2022.3198504

[21] Y.P. Deng, L. Wang, H. Jia, X.H. Zhang, Tong Xiangqian. A deep learning method based on bidirectional wavenet for voltage sag state estimation via limited monitors in power system. Energies, 2022, 15(6), 2273-2273. DOI: 10.3390/en15062273

[22] C. Galvez, A. Abur. Fault location in power networks using a sparse set of digital fault recorders. IEEE Transactions on Smart Grid, 2022, 13(5), 3468-3480. DOI: 10.1109/TSG.2022.3168904

[23] Y.Z. Chen, R.A. Jacob, Y.R. Gel, J. Zhang, H.V. Poor. Learning power grid outages with higher-order topological neural networks. IEEE Transactions on Power Systems, 2023, 39(1), 720-732. DOI: 10.1109/TPWRS.2023.3266956

[24] H.Y. Wui, M.H. Wang, Z. Xu, Y.W. Jia. Graph attention enabled convolutional network for distribution system probabilistic power flow. IEEE Transactions on Industry Applications, 2022, 58(6), 7068-7078. DOI: 10.1109/TIA.2022.3202159

[25] Y. Zhang, M. Yue, J.H. Wang, S. Yoo. Multi-agent graph-attention deep reinforcement learning for post-contingency grid emergency voltage control. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(3), 3340-3350. DOI: 10.1109/TNNLS.2023.3341334

[26] Z. Li, L. Ye, Y.N. Zhao, M. Pei, P. Lu, Y.L. Li, et al. A spatiotemporal directed graph convolution network for ultra-short-term wind power prediction. IEEE Transactions on Sustainable Energy, 2022, 14(1), 39-54. DOI: 10.1109/TSTE.2022.3198816

[27] X.H. Zhan, L. Kou, M.T. Xue, J.L. Zhang, L. Zhou. Reliable long-term energy load trend prediction model for smart grid using hierarchical decomposition self-attention network. IEEE Transactions on Reliability, 2022, 72(2), 609-621. DOI: 10.1109/TR.2022.3174093

[28] A. Ahmadi, M.C. Smith, E.R. Collins, V. Dargahi, S. Jin. Fast Newton-Raphson power flow analysis based on sparse techniques and parallel processing. IEEE Transactions on Power Systems, 2021, 37(3), 1695-1705. DOI: 10.1109/TPWRS.2021.3116182

[29] J.Y. Huang, L. Guan, Y.S. Su, H.C. Yao, M.X. Guo, Z. Zhong. System-scale-free transient contingency screening scheme based on steady-state information: A pooling-ensemble multi-graph learning approach. IEEE Transactions on Power Systems, 2021, 37(1), 294-305. DOI: 10.1109/TPWRS.2021.3097331

[30] J. Kwon, J. Kong, A. Munir. Sparse convolutional neural network acceleration with lossless input feature map compression for resource-constrained systems. IET Computers & Digital Techniques, 2022, 16(1), 29-43. DOI: 10.1049/cdt2.12038

[31] B. Najafi, S. Parsaeefard, A. Leon-Garcia. Entropy-aware time-varying graph neural networks with generalized temporal hawkes process: Dynamic link prediction in the presence of node addition and deletion. Machine Learning and Knowledge Extraction, 2023, 5(4), 1359-1381. DOI: 10.3390/make5040069

[32] H.H. Alhelou, N. Nagpal, H. Nagpal, P. Siano, M. AL-Numay. Dynamic state estimation for improving observation and resiliency of interconnected power systems. IEEE Transactions on Industry Applications, 2023, 60(2), 2366-2380. DOI: 10.1109/TIA.2023.3341060

[33] H.B. Zhang, W. Xiang, W.X. Lin, J.Y. Wen. Grid forming converters in renewable energy sources dominated power grid: Control strategy, stability, application, and challenges. Journal of Modern Power Systems and Clean Energy, 2021, 9(6), 1239-1256. DOI: 10.35833/MPCE.2021.000257

[34] M. Mansourlakouraj, H. Hosseinpour, H. Livani, M. Benidris. Waveform measurement unit-based fault location in distribution feeders via short-time matrix pencil method and graph neural network. IEEE Transactions on Industry Applications, 2022, 59(2), 2661-2670. DOI: 10.1109/TIA.2022.3231586

[35] X. Chen, G.N. Qu, Y.J. Tang, S. Low, N. Li. Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. IEEE Transactions on Smart Grid, 2022, 13(4), 2935-2958. DOI: 10.1109/TSG.2022.3154718

[36] R.I. Hamilton, P.N. Papadopoulos. Using SHAP values and machine learning to understand trends in the transient stability limit. IEEE Transactions on Power Systems, 2023, 39(1), 1384-1397. DOI: 10.1109/TPWRS.2023.3248941

[37] A. Aligholian, H. Mohsenian-Rad. GraphPMU: Event clustering via graph representation learning using locationally-scarce distribution-level fundamental and harmonic PMU measurements. IEEE Transactions on Smart Grid, 2022, 14(4), 2960-2972. DOI: 10.1109/TSG.2022.3225373

[38] W.Y. Guo, S.J. Liu, L.G. Weng, X.Y. Liang. Power Grid Load Forecasting Using a CNN-LSTM Network Based on a Multi-Modal Attention Mechanism. Applied Sciences, 2025, 15(5). DOI: 10.3390/app15052435

[39] X.X. Yu, Z.G. Zhang, B.P. Tang, M.H. Zhao. A multi-head self-attention autoencoder network for fault detection of wind turbine gearboxes under random loads.

Measurement Science and Technology, 2024, 35(8), 086137. DOI: 10.1088/1361-6501/ad4dd4

[40] H.C. Shu, H.M. Liu, Y.T. Tang, X. Su, Y.M. Han, Y. Dai. Fault Identification Method for Measured Travelling Wave of Transmission Line Based on CSCRFAM-Transformer. Protection and Control of Modern Power Systems, 2025, 10(2), 69-82. DOI: 10.23919/PCMP.2023.000322

[41] J. Luo, J.Y. Huang, J.C. Ma, S.Y. Liu. Application of self-attention conditional deep convolutional generative adversarial networks in the fault diagnosis of planetary gearboxes. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2024, 238(2), 260-273. DOI: 10.1177/1748006X221147784

[42] M.A. Hasnat, M. Rahnamay-Naeini. Detecting and locating cyber and physical stresses in smart grids using the k-nearest neighbour analysis of instantaneous correlation of states. IET Smart Grid, 2021, 4(3), 307-320. DOI: 10.1049/stg2.12030