Knowledge-Guided Deep Learning Scheme for Cross-System Adaptive Power Flow Analysis

Yongzhe Li
School of Electric Power
South China University of Technology
Guangzhou, China
1070671983@qq.com

Lin Guan*
School of Electric Power
South China University of Technology
Guangzhou, China
lguan@scut.edu.cn

Zihan Cai
School of Electric Power
South China University of Technology
Guangzhou, China
epc_zihan@mail.scut.edu.cn

***Abstract*—This paper proposes a Cross-System Adaptive Power Flow Analysis framework (CSA-PFA) based on the physical principles of power flow, aiming to enhance the adaptability of deep learning models across different power systems. Considering the local separability of power flow distributions, a local topology slicing method is developed to divide the power system into subgraphs with diverse structural characteristics for model training, guiding the model to learn scale-independent physical laws. Furthermore, a power flow diffusion-guided random walk (PDRW) module is designed to capture the intrinsic relationship between power flow allocation and branch admittance. By simulating the local power transfer process, PDRW strengthens the model's ability to recognize local topological and impedance differences among different systems. Experimental results on multiple standard and real-world systems demonstrate that the proposed CSA-PFA framework achieves superior cross-system adaptability and generalization performance across systems with various scales and topologies.**

Keywords—power flow analysis, deep learning, cross-system adaptability, local topology slicing, power flow diffusion-guided random walk

I. Introduction

In power systems with high renewable energy penetration, the strong uncertainty of renewable generation and the increasingly stringent *N-k* security and stability constraints lead to an exponential growth in the number of scenarios required for routine system operation analysis [1], imposing tremendous computational burdens on real-time power flow analysis (PFA). A well-trained deep learning (DL) model can directly map the initial power flow states to steady-state results during inference, achieving over 100× speedup compared with traditional numerical iterative methods [2]. Therefore, DL-based approaches can serve as a front-end surrogate for numerical solvers, handling the majority of operating scenarios and significantly reducing the computational load of online security and stability assessments [3].

The adaptability of DL models to different operating conditions has long been a critical issue. With the rapid progress of large language models (LLMs) and artificial intelligence for science, the development of a general and scalable PFA model has become an important research direction. Previous studies have shown that the prediction accuracy of DL-based PFA models usually increases with the model scale [4]. This suggests that a larger model, although requiring higher training cost, can achieve better performance. If a PFA model can be directly applied to, or slightly fine-tuned for, different systems, the training cost for cross-system deployment can be substantially reduced. Moreover, the power flow distribution of any power system is governed by physical constraints such as Kirchhoff's laws. This fundamental property provides a strong theoretical foundation for developing PFA models that are adaptive to different systems.

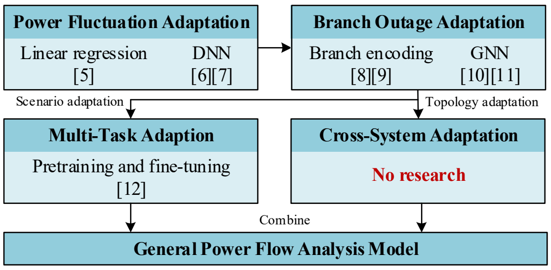

Fig. 1 Research landscape of DL-based PFA.

Fig. 1 illustrates the current research landscape of DL-based PFA models. Early studies mainly focused on adapting to variations in renewable generation and load under fixed network topologies. Linear regression [5] and deep neural networks (DNNs) [6][7] are among the first approaches that introduced AI techniques into PFA. Later, some studies improved the model's ability to identify topological changes by introducing additional encoding information, such as bus self-admittance [8] and branch outage indices [9]. Other studies adopt graph neural networks (GNNs), which naturally capture topological relationships in the grid [10][11]. Inspired by large-scale pretraining in LLMs, a pretraining and fine-tuning mechanism is introduced in [12]. It first pretrains a model on a large amount of unlabeled data to learn the fundamental patterns of power flow distribution, and then transfers it to multiple downstream tasks, such as power flow approximation and optimal power flow, with only a few samples. These studies have significantly enhanced the adaptability of DL-based PFA models. However, they are still limited to specific network scales. When the topology or scale of the system changes, new data collection and retraining are required. Enabling DL models to directly adapt to new systems would substantially improve their universality and reduce the cost of cross-system training.

The key to achieving such universality lies in joint training across multiple systems. Nevertheless, electrical characteristics differ notably among systems. These differences can be divided into global and local aspects. At the global level, system scale and connectivity vary considerably. For example, provincial transmission networks are generally designed with meshed or ring structures, whereas urban power networks often contain many radial configurations. At the local level, variations in branch impedance and subgraph topology cause distinct differences in power flow transmission paths and distribution patterns.

To address these challenges, this paper proposes a Cross-System Adaptive Power Flow Analysis framework (CSA-PFA). The proposed framework incorporates physical knowledge of power flow into both data processing and model

This work is supported by the Smart Grid-National Science and Technology Major Project of China (2025ZD0804900) and the National Natural Science Foundation of China (U22B6007).

979-8-3315-8138-1

design to enhance cross-system adaptability. The main contributions of this study are summarized as follows:

1) A CSA-PFA framework is developed to enhance DL model adaptability across systems by mitigating system differences and embedding physical priors.

2) A Local Topology Slicing (LTS) method is proposed, in which each local area of the system is treated as an independent sample. Subgraphs of different scales and connectivity are generated to reduce the impact of global structural differences.

3) A Power flow Diffusion-guided Random Walk (PDRW) method is developed to embed the relationship between branch impedance and power flow distribution into feature preprocessing. This improves the model's ability to capture and represent local topological differences and impedance characteristics.

II. Motivation and Framework Design

A. Key Challenges in Multi-System Joint Training

1) Differences in Global Connectivity

Different power systems vary significantly in both scale and connectivity. To quantitatively describe these global structural differences, we adopt two common graph-theoretic indicators from graph learning: the average node degree and the algebraic connectivity [13].

Let $\boldsymbol{X} \in \mathbb{R}^{N\times N}$ denote the bus reactance matrix, where N is the number of buses in the system. If a branch exists between buses i and j, the element X_{ij} represents the branch reactance; otherwise, $X_{ij} = 0$. Based on $\boldsymbol{X}$, the Laplacian matrix $\boldsymbol{L}$ can be constructed as $\boldsymbol{L} = \boldsymbol{D}^{-1/2}\boldsymbol{X}\boldsymbol{D}^{-1/2}$, where $\boldsymbol{D}$ is the degree matrix with diagonal elements $D_{ii} = \sum_j^N X_{ij}$.

The algebraic connectivity C is defined as the second smallest eigenvalue of $\boldsymbol{L}$, which reflects the connectivity strength of the system. The average node degree $\overline{D}$ can be calculated as:

$$\overline{D} = \sum_i^N \sum_j^N L_{ij} / N \qquad (1)$$

In this section, four standard systems (case36, case39, case57, and case118) and two real-world systems (rw1 and rw2) are analyzed. For each system, 2,000 samples are generated considering different branch maintenance conditions, and the corresponding connectivity strengths are calculated.

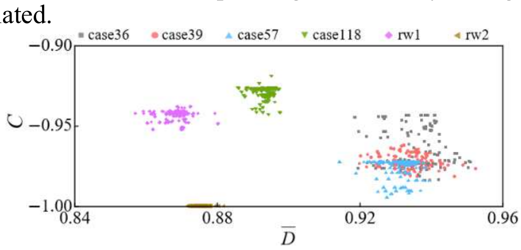

Fig. 2 Connectivity strength of different systems.

As illustrated in Fig. 2, the connectivity strength varies markedly across systems. If these samples are directly mixed for training, the model may tend to memorize the unique characteristics of each system, rather than learning the general physical laws that govern power flow distributions across systems.

2) Local Topological Differences

The feature extraction capability of DL models, especially GNNs, is strongly influenced by the local topology in which each bus is embedded. Differences in local connectivity and branch impedance directly affect how electrical quantities are transmitted within the network.

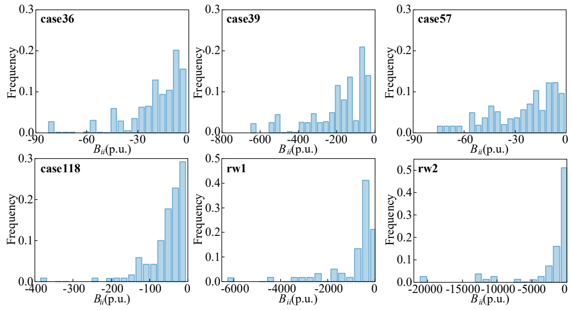

Fig. 3 Bus self-admittance distributions of different systems.

Fig. 3 visualizes the bus self-admittance distributions of the six systems. These distributions vary significantly among systems. This indicates that the relationships among buses and the impedance characteristics of branches differ widely, leading to distinct local power flow patterns and transmission behaviors.

If a DL model is trained directly on the raw input space, learning such complex local physical patterns becomes very difficult. In contrast, by introducing appropriate feature preprocessing to explicitly encode power flow transmission relationships into bus features, the model's sensitivity to local topological differences can be effectively reduced.

B. Design of CSA-PFA Framework

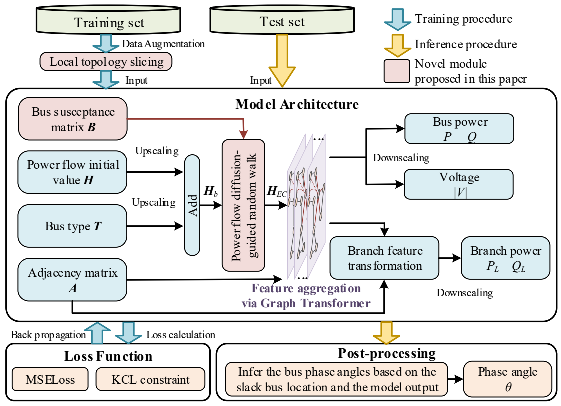

Fig. 4 Architecture of the CSA-PFA.

To address the aforementioned cross-system heterogeneity, a Cross-System Adaptive Power Flow Analysis (CSA-PFA) framework is proposed. The structure is illustrated in Fig. 4.

Since different systems have different slack buses and phase angle bases, it is difficult for the model to directly learn phase angle distributions. Therefore, both the input and output of the model exclude phase angle variables. The model input $\boldsymbol{H}$ includes the initial active power, reactive power, and voltage magnitude of all buses before power flow calculation. The model output consists of the complete post-calculation electrical quantities. To support branch-level security assessments, the model also outputs branch active power P_L and branch reactive power Q_L, which can be used for branch overload risk analysis and subsequent phase angle calculation. Besides, a bus-type matrix $\boldsymbol{T}$ is introduced to indicate whether each input element is known. An element of $\boldsymbol{T}$ equals 1 if the feature is known and 0 otherwise. The adjacency matrix $\boldsymbol{A}$ reflects the electrical coupling strength between buses and integrates information such as branch admittance and transformer tap ratios.

In the model structure, a Power flow Diffusion-guided Random Walk (PDRW) module is developed to enhance the recognition of local physical features. This method incorporates the idea of random walks and propagates bus features through their neighbors according to admittance relationships, thereby encoding local power flow transmission

979-8-3315-8138-1

behavior. Then, the processed features are fed into a feature aggregation network built with graph transformer layers [14] to learn the power flow correlations among buses. Finally, the complete high-dimensional representations are projected into lower-dimensional spaces to obtain the predicted results for buses and branches.

During model training, a Local Topology Slicing (LTS) strategy is introduced. The original system is decomposed into subgraphs of different scales and connectivity, which are used as training samples. This expansion increases the diversity of the training data and broadens the distribution of connectivity strength, helping the model learn generalizable power flow laws shared across different systems.

During inference, although CSA-PFA does not directly output phase angles, the model's outputs include voltage magnitudes V, branch active power P_L, and branch reactive power Q_L. The phase angle difference θ_{ij} between buses i and j can then be calculated using (2).

$$\theta_{ij} = \arctan \frac{b_{L,ij}P_{L,ij} + g_{L,ij}Q_{L,ij}}{g_{L,ij}P_{L,ij} - b_{L,ij}Q_{L,ij} - V_i^2(g_{L,ij}^2 + b_{L,ij}^2)} \tag{2}$$

where g_L+jb_L is the branch admittance between buses i and j.

By recursively computing these phase angle differences, the complete phase angle distribution of the system can be reconstructed.

III. Design of LTS

A. Physical Knowledge

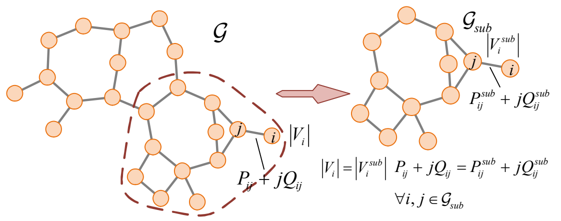

Fig. 5 Local topology slicing.

As illustrated in Fig. 5, a power system can be essentially regarded as a hierarchical network structure composed of multiple locally connected subgraphs that are coupled through tie lines. Because power flow obeys Kirchhoff's laws, the distribution of power naturally exhibits local separability. In other words, when a connected subgraph $\mathcal{G}_{sub} \subseteq \mathcal{G}$ is extracted from the complete network $\mathcal{G}$, the internal voltage and branch flow of $\mathcal{G}_{sub}$ remain unchanged if the power exchanged through tie lines with the external network is equivalently represented as loads.

Therefore, any local region within the power system can be treated as a sample for independent model training. This LTS method provides two major benefits. *First*, it significantly increases the number of training samples and enhances data diversity. *Second*, it expands the range of bus scales and wiring configurations within the dataset, preventing the model from memorizing system-specific features and thus improving its generalization and cross-system adaptability.

B. Methodology

Based on the above physical rationale, the proposed LTS procedure consists of six main steps. The process is described as follows.

1) Set the target number of buses in the subgraph as N_{sub}. Randomly select a bus b_0 from the power system as the starting point and initialize the subgraph bus set $\mathcal{V}_{sub} = \{b_0\}$.

2) For the current subgraph, search all buses directly connected to it and form the boundary bus set $\mathcal{B}$.

3) Randomly select a subset $\mathcal{B}_{sub}$ from the boundary set $\mathcal{B}$ that satisfies $0 < |\mathcal{B}_{sub}| \le \min\{|\mathcal{B}_{sub}|, N_{sub} - |\mathcal{V}_{sub}|\}$, and update the subgraph bus set as $\mathcal{B}_{sub} \leftarrow \mathcal{B}_{sub} \cup \mathcal{B}$.

4) Repeat 2) and 3) until the number of buses in the subgraph satisfies $|\mathcal{V}_{sub}| = N_{\text{sub}}$.

5) To preserve the original power flow characteristics after subgraph extraction, convert the power of all tie lines connecting the subgraph and the remaining system into equivalent loads. This ensures that the internal power balance and flow distribution of the subgraph remain consistent with those in the original network.

6) In the equivalent subgraph, introduce random line outages and stochastic variations in bus power injections to generate the final set of subgraph samples.

Through the LTS process, the original network is decomposed into a large number of representative and independent local subgraphs that preserve the physical consistency of local power flow distributions while varying in size and connectivity. By broadening the topological pattern distribution of the training set, LTS enhances the model's adaptability to diverse network structures and its generalization to unseen systems. Moreover, the increased structural diversity discourages system-specific memorization and promotes the learning of scale-independent power flow laws, thereby effectively mitigating overfitting during training.

IV. Design of PDRW

A. Physical Knowledge

In power systems, the power balance at each bus must satisfy Kirchhoff's Current Law (KCL), which requires that the total injected power equals the total outgoing power. Let S_i denote the injected complex power at bus i, and let $\mathcal{N}(i)$ represent its neighboring bus set. The complex power flowing through branch i-j is represented by $S_{L,ij}$. Then the KCL at bus i can be expressed as:

$$S_i = \sum_{j\in\mathcal{N}(i)} S_{L,ij} = \sum_{j\in\mathcal{N}(i)} V_i\left((V_i - V_j)Y_{ij} + V_i b_{m,ij}\right)^* \tag{3}$$

where Y_{ij} is the admittance between buses i and j, $b_{m,ij}$ is the shunt susceptance, and the superscript * represents the complex conjugate operator.

It is shown in (3) that the power distribution at bus i is closely related to the admittance of its connected branches. Since power flow tends to propagate along paths with larger admittances, the admittance distribution fundamentally determines the local flow characteristics of the system.

Based on this physical property, if the correlation between power flow distribution and branch admittance is explicitly encoded in the model, the difficulty of learning local flow behaviors can be greatly reduced, thereby improving the model's adaptability to different system structures.

B. Methodology

For transmission networks, branch resistance is usually much smaller than reactance, only the susceptive part of the branch admittance is considered in the encoding process. The bus susceptance matrix $\boldsymbol{B}$ is used as the foundation of the encoding.

Since the flow distribution is positively correlated with branch admittance, this rule aligns well with the principle of the random walk algorithm, in which the strength of bus connections determines the probability of information propagation. By analogy, the random walk matrix for power flow $\boldsymbol{R}_w$ is defined as:

$$\boldsymbol{R}_w = \boldsymbol{B}\boldsymbol{D}_B^{-1} \tag{4}$$

where $\boldsymbol{D}_B$ is the degree matrix of the susceptance matrix $\boldsymbol{B}$. Each element $R_{w,ij}$ represents the probability that power at bus i flows toward bus j.

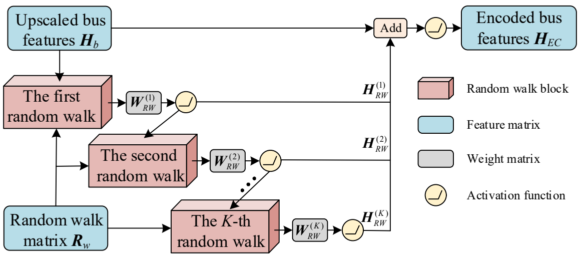

Fig. 6 Process of PDRW.

Based on the random walk matrix, the PDRW process is illustrated in Fig. 6. The method simulates the diffusion mechanism of power transmission in the network, allowing bus features to propagate hierarchically through multiple layers. During each random walk, bus i spreads its information to the neighboring buses according to the probabilities defined by $\boldsymbol{R}_w$, and the features of the neighboring buses are updated accordingly. Let $\boldsymbol{H}_{RW}^{(k)}$ denote the bus feature matrices after the k-th random walk with $\boldsymbol{H}_{RW}^{(0)} = \boldsymbol{H}_b$. The update rule is formulated as:

$$\boldsymbol{H}_{RW}^{(k+1)} = \sigma\left(\left(\boldsymbol{R}_w^{\mathrm{T}} \boldsymbol{H}_{RW}^{(k)}\right) \boldsymbol{W}_{RW}^{(k)}\right), k \geq 1 \tag{5}$$

where $\boldsymbol{W}_{RW}^{(k)}$ is the trainable weight matrix of the k-th walk, and $\sigma(\cdot)$ denotes the activation function.

After K random walks, the encoded bus representation $\boldsymbol{H}_{EC}$ is obtained as:

$$\boldsymbol{H}_{EC} = \sigma\left(\sum_{k=0}^{K} \boldsymbol{H}_{RW}^{(k)}\right) \tag{6}$$

Through PDRW, each bus feature integrates both local topological connectivity and branch admittance information. This enables the model to capture the intrinsic laws of power flow distribution across diverse systems, thereby improving its generalization capability in cross-system training.

Unlike standard GNNs, PDRW employs physically guided propagation weights that explicitly reflect power flow transmission tendencies between interconnected buses. Its diffusion mechanism is designed to simulate the actual transmission and allocation of power flow, rather than performing feature aggregation for representation learning. By embedding power flow propagation priors before deep feature extraction, PDRW significantly facilitates multi-system joint training.

V. Case Study

Six representative power systems are selected to verify the effectiveness of the proposed method, including four standard test systems (case36, case39, case57, case118) and two real-world systems (rw1 and rw2). System rw1 represents a provincial grid in southern China, and rw2 corresponds to an urban grid. The maximum number of buses in all samples is 120. To evaluate model generalization, case36 is excluded from training and used only for testing.

Based on the LTS method, 100,000 samples are generated for each of the five systems except case36, resulting in 500,000 training samples in total. In addition, considering 20% power fluctuations and 0-2 branch outages, 20,000 samples are generated per system for model testing.

The model is implemented using the MindSpore framework. After grid search tuning, the number of PDRW random walk layers is set to K = 3. Prediction accuracy is measured by the Normalized Root Mean Square Error (NRMSE) [11]:

$$\mathrm{NRMSE}(\boldsymbol{Y},\widetilde{\boldsymbol{Y}}) = \frac{1}{F}\sum_{j=1}^{F}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{(Y_{ij}-\widetilde{Y}_{ij})^2}{\mathrm{var}(\boldsymbol{Y}_j)}} \tag{7}$$

where F represents the electrical features involved in the calculation, including bus voltage magnitude, phase angle, and branch active and reactive power, while var(·) denotes the variance of each corresponding feature.

A. Model Performance on Different Systems

To verify the effectiveness of each module, three comparison models are designed. In M1, the LTS method is not used, and the training set is expanded to 500,000 samples only by introducing power fluctuations and partial branch outages. In M2, the PDRW module is removed. In M3, both LTS and PDRW are removed, and the training samples are generated in the same way as in M1. In M4 and M5, the PDRW is replaced with a graph convolutional network (GCN) and a graph attention network (GAT).

TABLE I NRMSE OF THE MODEL ON DIFFERENT SYSTEMS

Model	Seen					Unseen
	case39	case57	case118	rw1	rw2	case36
M1	1.426	2.957	1.320	0.854	3.979	54.385
M2	1.646	4.744	2.636	1.733	3.094	29.122
M3	2.181	4.862	3.320	2.337	4.228	69.619
M4	1.348	3.021	1.961	1.336	1.998	28.392
M5	1.672	3.503	2.566	1.774	2.693	29.006
Proposed	**0.977**	**2.184**	**1.026**	**0.614**	**1.710**	**26.673**

All the data in all tables are scaled by 100 times the actual NRMSE values.

Table I lists the NRMSE results for all systems. The results show that both LTS and PDRW significantly improve model accuracy and generalization. With LTS, the average error decreases by 33.0% on seen systems and 51.0% on unseen systems. With PDRW, the error decreases by 53.0% and 8.4%, respectively. When both modules are used together, the error reduction reaches 62.5% and 61.7%.

When PDRW is replaced by standard GNN modules, the performance is significantly inferior to that of the proposed PDRW-based model. This is because standard GNNs rely heavily on bus and branch attributes during feature aggregation and fail to achieve a unified modeling of power flow transmission laws across different systems, which limits their cross-system adaptability.

To further evaluate performance, CSA-PFA is compared with the DC Power Flow (DCPF) method, which only predicts active power on branches. For fairness, NRMSE is calculated using only the branch active power. As shown in Table II, CSA-PFA achieves much higher accuracy than DCPF on seen systems. On the unseen case36 system, its performance is comparable to DCPF, demonstrating strong cross-system generalization.

TABLE II COMPARISON RESULTS WITH DCPF

Model	Seen					Unseen
	case39	case57	case118	rw1	rw2	case36
DCPF	5.989	13.615	29.221	24.654	2.696	6.726
Proposed	0.342	1.063	1.002	0.584	0.226	5.161

B. Advantages of Multi-System Joint Training

To evaluate the benefit of joint training, we compare single-system training and multi-system joint training. In the single-system setting, the model is trained only with data from one system and tested on both its own test set and case36.

Table III summarizes the results. Joint training significantly improves prediction accuracy for all systems. Because the model is exposed to samples from multiple scales and topologies, it learns more general power flow laws, which greatly enhance generalization to the unseen case36 system.

TABLE III COMPARISON BETWEEN JOINT TRAINING AND SINGLE-SYSTEM TRAINING

Training strategy		case39	case57	case118	rw1	rw2
Single-system training	Test set	1.970	3.898	1.822	1.058	2.933
	case36	58.589	234.326	377.385	78.746	55.692
Joint training	Test set	0.977	2.184	1.026	0.614	1.710
	case36	26.673				

To further validate the generality of the proposed cross-system learning framework. Specifically, one system is alternately excluded from training among the six systems, while the remaining five systems are used for joint training. Both the jointly trained model (JT) and the corresponding single-system trained model (ST) are then evaluated on the excluded system. The average prediction errors are summarized in Table IV. The results show that cross-system joint training consistently improves prediction accuracy across different systems by at least 61%, demonstrating the general and robust effectiveness of the proposed framework for cross-system generalization.

TABLE IV GENERALIZATION VALIDATION WITH DIFFERENT SYSTEMS HELD OUT

Excluded system	case36	case39	case57	case118	rw1	rw2
ST	160.95	171.36	193.00	118.50	158.78	162.81
JT	26.67	36.83	44.10	46.08	32.61	20.50

C. *Effect of Local Topology Slicing*

To visualize the impact of LTS on data distribution, 2,000 samples are randomly selected from each system with and without LTS. The connectivity strength of these samples is compared with that of the case36 test set, as shown in Fig. 7.

When LTS is not used, the training set distribution shows a clear deviation from the case36 test set. After applying LTS, the distribution range becomes much wider and covers the unseen system more comprehensively. This demonstrates that LTS effectively reduces the distribution gap between training and new systems, thereby improving the model's performance on case36.

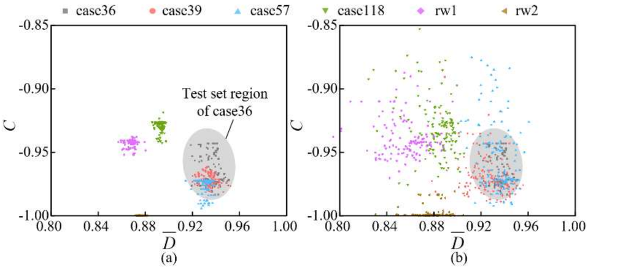

Fig. 7 Distribution of connectivity strength between training and case36 test sets. (a) Without LTS. (b) With LTS.

D. *Effect of Power Flow Diffusion-Guided Random Walk*

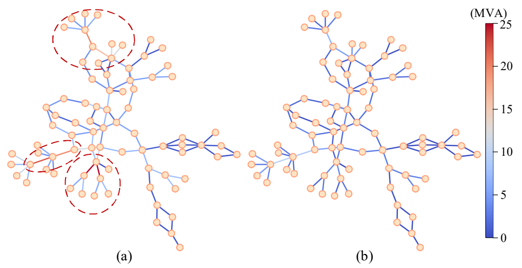

Fig. 8 Branch power prediction error in rw2. (a) M2. (b) Proposed.

To analyze the contribution of PDRW, Fig. 8 shows the 90th-percentile prediction error of each branch in rw2, with and without PDRW. The rw2 is a typical urban system that contains many radial structures, which appear less frequently in the training data. In Fig. 8(a), when PDRW is removed, the prediction error in these radial areas increases noticeably. In contrast, after PDRW is applied, the model learns local transmission patterns through the random walk mechanism and accurately captures admittance-based relationships among branches. This results in significantly improved prediction accuracy in rw2.

CONCLUSION

This paper proposes a CSA-PFA framework that integrates physical principles into model design and training to improve the universality and generalization of DL models across different power systems. The LTS method expands the training data distribution by slicing the network into local subgraphs, while PDRW simulates power flow diffusion through random walks to enhance local feature representation. Experimental results show that CSA-PFA achieves superior prediction accuracy and strong generalization, with average errors reduced by 62.53% on seen systems and 61.69% on unseen systems.

For distribution networks or strongly converter-dominated systems, branch resistance is often non-negligible, and the current PDRW formulation may require further refinement. Future work will extend the diffusion design of PDRW to improve its applicability to such systems, for example, by adopting a multi-channel propagation mechanism that jointly incorporates conductance and reactance into feature diffusion.

REFERENCES

[1] Y. Li, Y. Xu, W. Gu, S. Lu and L. Mili, "Global sensitivity analysis for rare events in probabilistic power flow using subset simulation," *IEEE Trans. Power Syst.*, vol. 40, no. 2, pp. 1969-1972, Mar. 2025.

[2] Y. Du, F. Li, J. Li and T. Zheng, "Achieving 100x acceleration for N-1 contingency screening with uncertain scenarios using deep convolutional neural network," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3303-3305, Jul. 2019.

[3] Y. Li, et al., "Data-physics joint-driven N-k static security assessment with higher-order graph convolution model," *Int. J. Elect. Power Energy Syst.*, vol. 172, p. 111128, Nov. 2025.

[4] Y. Zhu, Y. Zhou, L. Yan, Z. Li, H. Xin and W. Wei, "Scaling graph neural networks for large-scale power systems analysis: empirical laws for emergent abilities," *IEEE Trans. Power Syst.*, vol. 39, no. 6, pp. 7445-7448, Nov. 2024.

[5] Y. Chen, C. Wu and J. Qi, "Data-driven power flow method based on exact linear regression equations," *J. Mod. Power Syst. Clean Energy*, vol. 10, no. 3, pp. 800-804, May 2022.

[6] X. Hu, H. Hu, S. Verma and Z.-L. Zhang, "Physics-guided deep neural networks for power flow analysis," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2082-2092, May 2021.

[7] Y. Yang, Z. Yang, J. Yu, B. Zhang, Y. Zhang and H. Yu, "Fast calculation of probabilistic power flow: a model-based deep learning approach," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2235-2244, May 2020.

[8] B. Tan, J. Zhao and Y. Chen, "Scalable risk assessment of rare events in power systems with uncertain wind generation and loads," *IEEE Trans. Power Syst.*, vol. 40, no. 2, pp. 1374-1388, Mar. 2025.

[9] M. Xiang, J. Yu, Z. Yang, Y. Yang, H. Yu and H. He, "Probabilistic power flow with topology changes based on deep neural network," *Int. J. Elect. Power Energy Syst.*, vol. 117, p. 105650, May 2020.

[10] M. Gao, J. Yu, Z. Yang and J. Zhao, "Physics embedded graph convolution neural network for power flow calculation considering uncertain injections and topology," *IEEE Trans. on Neural Netw. Learn. Syst.*, vol. 35, no. 11, pp. 15467-15478, Nov. 2024.

[11] J. B. Hansen, S. N. Anfinsen and F. M. Bianchi, "Power flow balancing with decentralized graph neural networks," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2423-2433, May 2023.

[12] Y. Zhu, Y. Zhou, W. Wei, P. Li and W. Huang, "GNNs' generalization improvement for large-scale power system analysis based on physics-informed self-supervised pre-training," *IEEE Trans. Power Syst.*, vol. 40, no. 5, pp. 4145-4157, Sept. 2025.

[13] S. Kirkland and S. Fallat, "Perron components and algebraic connectivity for weighted graphs," *Linear Multilinear Algebra*, vol. 44, no. 2, pp. 131-148, 1998.

[14] Y. Li et al., "A multi-task power flow analysis model for power flow calculation and convergence enhancement based on parallel attention," *CSEE J. Power Energy Syst.*, Jan. 2026, doi: 10.17775/CSEEJPES.2025.00840.

979-8-3315-8138-1