
Graph Neural Networks for Learning Real-Time Prices in Electricity Market

Shaohui Liu¹ Chengyang Wu¹ Hao Zhu¹

Abstract

Solving the optimal power flow (OPF) problem in real-time electricity market improves the efficiency and reliability in the integration of low-carbon energy resources into the power grids. To address the scalability and adaptivity issues of existing end-to-end OPF learning solutions, we propose a new graph neural network (GNN) framework for predicting the electricity market prices from solving OPFs. The proposed GNN-for-OPF framework innovatively exploits the locality property of prices and introduces physics-aware regularization, while attaining reduced model complexity and fast adaptivity to varying grid topology. Numerical tests have validated the learning efficiency and adaptivity improvements of our proposed method over existing approaches.

1. Introduction

Electricity market pricing is one of the most crucial tasks of operating large-scale power grids. As part of the deregulated electricity market, real-time market determines the incremental adjustment to the day-ahead dispatch by solving the optimal power flow (OPF) problem (Cain et al., 2012), which aims at the most economic decisions for the flexible generation or demand while satisfying a variety of safety-related network constraints. The real-time OPF or market pricing is instrumental for ensuring high efficiency and reliability of grid operations (Cain et al., 2012), particularly under the increasing integration of intermittent and variable resources towards a low-carbon energy future.

Energy sources to a majority of electric power grids rely heavily on fossil fuels, which contribute to a significant portion of world-wide greenhouse gas (GHG) emissions. Such environment effects can be directly accounted by the formulation of OPF to reduce GHG emissions (Yang et al.,

2015; Li et al., 2011; Shao & Jewell, 2010; Gholami et al., 2014). Meanwhile, the increasing integration of renewable sources such as wind and solar can greatly diminish the overall carbon footprint of electricity generation. However, the variability of renewable generation poses as a main challenge in real-time solutions of OPF problem. As a result, the grid operating conditions change rapidly, traditional iterative methods initialized by the past solutions (Castillo & O’Neill, 2013) may incur convergence and computation issues.

The accurate ac-OPF problem is known to incur high computation complexity due to its non-linear, non-convex formulation (Cain et al., 2012). For efficient online solution, machine learning (ML) techniques have been recently advocated through extensive off-line training of neural network (NN) models. Existing ML-for-OPF approaches have focused on identifying the active constraints (Misra et al., 2018; Deka & Misra, 2019; Chen & Zhang, 2020), finding a warm start for iterative OPF solutions (Baker, 2019), or addressing the feasibility issue (Pan et al., 2019; Guha et al., 2019; Zamzam & Baker, 2020). Almost all of them rely on *end-to-end* NNs, which incur high model and computation complexity for large-scale power grids. In addition to scalability issue, they need to be constantly re-trained whenever the system inputs change as a result of frequently varying grid resources or topology. Thus, existing approaches fall short in efficiently transferring the knowledge obtained from off-line training into fast, adaptive online OPF decisions.

To tackle these challenges, we propose to leverage the graph neural networks (GNNs) to design a topology-aware OPF learning framework. The GNN architecture (Kipf & Welling, 2016; Gama et al., 2020; Garg et al., 2020) can effectively incorporate graph-based embedding of nodal features and explore the topology structures of the underlying prediction models. While a very recent work (Owerko et al., 2020) has used GNNs to predict OPF’s nodal power injections, the latter mainly depends on the cost of dispatching each resource and does not share any topology-based similarity, or the *locality property* that is ideal for GNN-based predictions. Hence, we instead advocate to predict the actual OPF outputs for electricity market, namely the locational marginal prices (LMPs) known as the real-time market signals (Wood et al., 2013). As LMPs relate to the duality analysis for OPF, their dependence on grid topology

¹Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, United States. Correspondence to: Shaohui Liu <shaohui.liu@utexas.edu>, Hao Zhu <haozhu@utexas.edu>.

has been recognized in (Jia et al., 2013; Geng & Xie, 2016).

To this end, we have introduced the ac- and dc-OPF problem formulations (Section 2) and exploited the topological structure of LMPs to design a GNN-for-OPF learning framework (Section 3). This physics-aware approach not only capitalizes on the locality property of LMPs, but also motivates meaningful regularization on the feasibility of OPF line limits. Numerical results (Section 4) have demonstrated the high prediction performance of the proposed GNN-for-LMP approach at reduced model complexity, while confirming its topology adaptivity as an effective transfer learning tool to deal with fast varying grid topology in real-time markets.

2. Real-Time Market Modeling

Consider a power grid modeled by an undirected graph $G = (\mathcal{V}, \mathcal{E})$. The node set $\mathcal{V}$ consists of N nodes, each connected to loads or generators, while the edge set $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ includes transmission lines or transformers. Let $\mathbf{p}, \mathbf{q} \in \mathbb{R}^N$ collect the nodal active and reactive power injections, respectively; and similarly for voltage $\mathbf{v} \in \mathbb{R}^N$. Given the network admittance (Y-bus) matrix $\mathbf{Y} \in \mathbb{R}^{N \times N}$, the ac-OPF problem is formulated as

$$\min_{\mathbf{p}, \mathbf{q}, \mathbf{v}} \sum_{i=1}^N c_i(p_i) \quad (1a)$$

$$\text{s.t. } \mathbf{p} + \mathbf{j}\mathbf{q} = \text{diag}(\mathbf{v})(\mathbf{Y}\mathbf{v})^* \quad (1b)$$

$$\underline{\mathbf{V}} \leq |\mathbf{v}| \leq \bar{\mathbf{V}} \quad (1c)$$

$$\underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}} \quad (1d)$$

$$\underline{\mathbf{q}} \leq \mathbf{q} \leq \bar{\mathbf{q}} \quad (1e)$$

$$\underline{f}_{ij} \leq f_{ij}(\mathbf{v}) \leq \bar{f}_{ij}, \quad \forall (i, j) \in \mathcal{E} \quad (1f)$$

where $c_i(\cdot)$ is a convex (typically quadratic or piece-wise linear) cost function for flexible nodal injections. The equality (1b) ensures nodal power balance, while constraints (1c)-(1f) list the various operational limits such as line flow limits in (1f). This general OPF (1) includes both flexible generation and demand, with negative injections indicating the latter.

To simplify the nonlinear, non-convex problem (1), the linear dc-OPF is widely used for solving $\mathbf{p}$ only, as

$$\min_{\mathbf{p}} \sum_{i=1}^N c_i(p_i) \quad (2a)$$

$$\text{s.t. } \mathbf{1}^\top \mathbf{p} = 0 \quad (2b)$$

$$\underline{\mathbf{p}} \leq \mathbf{p} \leq \bar{\mathbf{p}} \quad (2c)$$

$$\underline{\mathbf{f}} \leq \mathbf{S}\mathbf{p} \leq \bar{\mathbf{f}} \quad (2d)$$

where matrix $\mathbf{S}$ is the injection shift factor (ISF) matrix to form the line flow $\mathbf{f} = \mathbf{S}\mathbf{p}$ with the limit $\underline{\mathbf{f}} = -\bar{\mathbf{f}}$. Compared to (1), the dc-OPF problem omits the modeling of reactive power and voltage, and also uses lossless linearized power flow to simplify power balance as in (2b). The accuracy of

dc-OPF can be improved by considering better linearization around the operating points and including line losses; see e.g., (Garcia, 2019). As the resultant constraints are still linear, the generalized dc-OPF problem can be easily computed using off-the-shelf convex solvers.

Learning the OPF solutions amounts to obtaining the mapping from the uncontrolled problem inputs to the OPF outputs. In real-time market (1)-(2), nodal injections have uncontrollable components $\mathbf{p}^u$ and $\mathbf{q}^u$ from variable demand or renewable resources. They in turn affect the limits of respective injections in (1)-(2). In addition, the cost function $c_i(\cdot)$ depends on the offers submitted by generation or load serving entities (LSEs), thus varying for each OPF instance as well. Hence, for each node i the input variables include $\mathbf{x}_i \triangleq [\bar{p}_i, \underline{p}_i, \bar{q}_i, \underline{q}_i, \mathbf{c}_i] \in \mathbb{R}^d$, with $\mathbf{c}_i$ denoting the $(d-4)$ parameters used for defining the nodal cost function. For example, quadratic cost is given by the quadratic and linear coefficients, while piece-wise linear one by the change points and gradients of each linear part. Due to increasing variability of resources and offers, the real-time OPF problems may experience dramatic changes from instance to instance. Given this vast variability, it is beneficial to develop a learning-based approach that can enable efficient real-time market operations.

3. Topology-aware Learning for Market Prices

We advocate a topology-aware graph neural network (GNN) based framework for learning real-time prices that attains high learning efficiency and topology adaptivity. Before introducing GNNs, we first discuss how locational marginal prices (LMPs), the outputs of OPF, are connected to the grid topology G . LMPs are market signals used by each generator or demand to determine the flexible power injection in order to minimize its own cost. To show the topology dependence, consider the simple convex dc-OPF problem (2), for which dual variables λ , and $[\underline{\mu}; \bar{\mu}]$ are introduced for constraints (2b) and (2d), respectively, with (2c) kept as an implicit constraint. Given the optimal dual variables (denoted by $*$), the nodal LMP vector is given by

$$\boldsymbol{\pi} \triangleq \lambda^* \cdot \mathbf{1} - \mathbf{S}^\top (\bar{\boldsymbol{\mu}}^* - \underline{\boldsymbol{\mu}}^*) \quad (3)$$

using the ISF matrix $\mathbf{S}$. Interestingly, vector $(\bar{\boldsymbol{\mu}}^* - \underline{\boldsymbol{\mu}}^*)$ indicates the congested lines due to complimentary slackness (Boyd et al., 2004); i.e., $\bar{\mu}_\ell^* - \underline{\mu}_\ell^* = 0$ if and only if line ℓ is reaching limit $\bar{f}_\ell$ ($\underline{f}_\ell$). Clearly, the LMP $\boldsymbol{\pi}$ only depends on those congested lines that have non-zero $(\bar{\mu}_\ell^* - \underline{\mu}_\ell^*)$. Interestingly, matrix $\mathbf{S}$ strongly depends on the graph topology such that $\boldsymbol{\pi}$ has the *locality* property that is perfectly suited for GNNs. Typically, only a few transmission lines are actually congested (Price & Goodin, 2011). Thus, LMPs tend to be similar within the neighboring nodes. Formally, matrix $\mathbf{S}$ depends on graph incidence matrix $\mathbf{A}_r$ and a di-

agonal matrix with line reactance values $\mathbf{X} = \text{diag}\{x_{ij}\}$, as well as the resultant weighted graph Laplacian matrix $\mathbf{B}_r = \mathbf{A}_r^\top \mathbf{X}^{-1} \mathbf{A}_r$. Both $\mathbf{A}_r$ and $\mathbf{B}_r$ are reduced from the original matrices by eliminating a reference node to obtain the full-rank counterparts. Given the compact singular value decomposition (SVD) $\mathbf{A}_r^\top \mathbf{X}^{-\frac{1}{2}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$, we can write the ISF matrix as

$$\mathbf{S}^\top = \mathbf{B}_r^{-1} \mathbf{A}_r^\top \mathbf{X}^{-1} = \mathbf{U} \mathbf{\Sigma}^{-1} \mathbf{V}^\top \mathbf{X}^{-\frac{1}{2}} \quad (4)$$

with the eigen-decomposition $\mathbf{B}_r = \mathbf{U} \mathbf{\Sigma}^2 \mathbf{U}^\top$. Thus, the LMP vector $\boldsymbol{\pi}$ in (3) is exactly generated by the eigenspace of the Laplacian $\mathbf{B}_r$, which can be viewed as a graph shift operator (GSO) (Ramakrishna & Scaglione, 2021). Accordingly, it strongly depends on the graph topology, which motivates one to use the topology-aware GNN models for prediction. Note that even though this LMP analysis corresponds to the simple dc-OPF, similar intuitions also hold for the ac-OPF problem; see e.g., (Garcia, 2019).

In the OPF problem, we aim to obtain the function $f(\mathbf{X}) \rightarrow \boldsymbol{\pi}$, where the input $\mathbf{X} \in \mathbb{R}^{N \times d}$ has the nodal features $\{\mathbf{x}_i\}$ as its rows. To model $f(\cdot)$ using fully-connected NN (FCNN), the input to first layer $\mathbf{X}_0$ can be a vector embedding of $\mathbf{X}$, with each layer t as

$$\mathbf{X}_{t+1} = \sigma(\mathbf{W}_t \mathbf{X}_t + \mathbf{b}_t), \quad \forall t = 0, \dots, T-1 \quad (5)$$

where $\mathbf{W}_t$ and $\mathbf{b}_t$ are parameters to be learned, while $\sigma(\cdot)$ is the nonlinear activation like ReLU. Albeit generalizable to a variety of *end-to-end* learning tasks, the FCNN models would incur significant scalability issue for large-scale OPF learning. It is possible to reduce the layer complexity by using the graph topology, leading to graph-pruned NNs. For example, the *graph-induced deep NN* (GiDNN) developed in (Zamzam & Sidiropoulos, 2020) sparsifies matrix $\mathbf{W}_t$ according to the graph topology. By pruning out a majority of blocks in $\mathbf{W}_t$, the total number of parameters is reduced.

Inspired by the graph signal viewpoint on $\boldsymbol{\pi}$ arising from the structure in (4), we propose to systematically reduce the prediction model complexity by leveraging the GNN architecture (Isufi et al., 2020; Ma & Tang, 2020; Kipf & Welling, 2016). As a special case of NNs, GNNs take the input features $\{\mathbf{x}_i\}$ defined over graph nodes in $\mathcal{V}$, with each layer aggregating only the embedding of neighboring nodes. In this sense, GNN is ideal for predicting output labels having locality property as a result of graph diffusion processes. To define the GNN layers, consider again the feature matrix $\mathbf{X}$ as the input to the first layer $\mathbf{X}_0$, and each layer t now becomes:

$$\mathbf{X}_{t+1} = \sigma(\mathbf{W} \mathbf{X}_t \mathbf{H}_t + \mathbf{b}_t), \quad \forall t = 0, \dots, T-1 \quad (6)$$

where the feature filters $\{\mathbf{H}_t\}$ are the $(d_t \times d_{t+1})$ parameter matrices that are learned through training, which do not

change with system size N . The key of GNNs lies in the graph convolution filter $\mathbf{W} \in \mathbb{R}^{N \times N}$ such that the node embedding is updated by neighborhood aggregation. Matrix $\mathbf{W}$ can be the (weighted) graph Laplacian or adjacency matrix, or its normalized version for stability concerns (Isufi et al., 2020). For better performance, it can also be learned through training, leading to a bi-linear filtering process in (6) as developed in (Isufi et al., 2020). In this case, $\mathbf{W}$ has the sparsity structure as the graph Laplacian with number of non-zero parameters proportional to the weight of edges. Clearly, the GNN architecture can significantly reduce the number of parameters per layer. As the average node degree of real-world power grids is around 2 or 3 (Birchfield et al., 2016), we have the following result.

AS1. *The edges are very sparse, and the number of edges $|E| \sim \mathcal{O}(|V|) = \mathcal{O}(N)$.*

Proposition 1. *Under (AS1) and by defining $D = \max_t \{d_t\}$, the number of parameters for each bi-linear GNN layer in (6) is $\mathcal{O}(N + D^2)$.*

This complexity order result follows easily from checking the number of nonzero entries in $\mathbf{W}$ and $\mathbf{H}_t$ in (6). Trainable graph filter $\mathbf{W}$ only increases the complexity by the number of edges, which scales linearly with N thanks to (AS1). Compared to $\mathcal{O}(N^2 D^2)$ as the number of parameters in each FCNN layer, the GNN architecture scales very gracefully with the network dimension. Thanks to the locality property of LMP $\boldsymbol{\pi}$, our proposed design can greatly improve computation time and generalization performance by utilizing the reduced-complexity GNN models.

Feasibility-based Regularization: As OPF is a network-constrained problem, we design the loss function for learning LMPs that can account for the solution feasibility and constraints. Note that for dc-OPF problem, the LMP fully determines the decision variables in $\mathbf{p}$. Based on the KKT optimality condition (Boyd et al., 2004), the predicted LMP $\hat{\boldsymbol{\pi}}$ allows to obtain the optimal nodal injection, as

$$p_i^* = \arg \min_{p_i \leq p_i \leq \bar{p}_i} c_i(p_i) - \hat{\pi}_i p_i, \quad \forall i \in \mathcal{V} \quad (7)$$

For quadratic (or generally strongly-convex) cost functions, the solution is unique by comparing the unconstrained minimum with the boundary points $[p_i, \bar{p}_i]$. As for (piece-wise) linear cost functions, this also holds for most nodes if the derivative $c'_i(p_i^*) \neq \hat{\pi}_i$. Otherwise, the optimal p_i^* at the other nodes can still be computed from the power balance of the full system and congested lines.

Using this result, we advocate the following chain to generate the corresponding nodal injection and line flow solutions to the predicted LMP $\hat{\boldsymbol{\pi}}$:

$$\mathbf{X} \xrightarrow{f(\mathbf{X}; \boldsymbol{\theta})} \hat{\boldsymbol{\pi}} \xrightarrow{(7)} \hat{\mathbf{p}}^*(\hat{\boldsymbol{\pi}}) \xrightarrow{\mathbf{S}} \hat{\mathbf{f}}^*(\hat{\boldsymbol{\pi}})$$

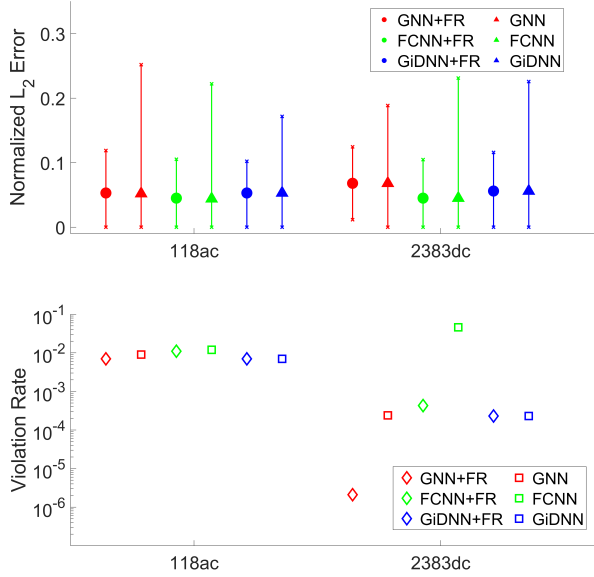


Figure 1. Comparison of GNN, FCNN, and GiDNN models, for both MSE loss and the feasibility regularized (FR) one, in terms of (top) the normalized L_2 error in predicting π (mean $\pm$ standard deviation); and (bottom) the violation rates of line limits (total violation level versus total limit) for the feasibility performance.

where $f(\cdot; \theta)$ denotes the GNN model with weight parameter θ according to (6). Hence, the predicted $\hat{\mathbf{p}}^*$ is strictly feasible for (2c), while the predicted $\hat{\mathbf{f}}^*$ can be used to regularize the GNN loss function by enforcing the feasibility of (2d). This way, the loss function for GNN training becomes

$$\mathcal{L}(\theta) := \|\pi - \hat{\pi}\|_2^2 + \lambda \|\sigma(|\hat{\mathbf{f}}^*(\hat{\pi})| - \bar{\mathbf{f}})\|_1 \quad (8)$$

where the second term captures the total line flow violation of the limit $\bar{\mathbf{f}}$, leading to LMP prediction more amendable to feasibility. Additional regularization terms on $\hat{\mathbf{p}}$ can be introduced such as the infinity error norm in predicting π .

4. Numerical Results

This section presents the efficiency and scalability results for the proposed GNN-based algorithms by using the 118- and 2383-bus systems from the IEEE PES PGLib-OPF benchmark library (Babaeinejadsarookolae et al., 2019). A small example on topology adaptivity is also included to demonstrate the proposed GNN models can quickly adapt to varying grid topology in real-time operations. We generated the datasets from solving the ac/dc-OPF problems for each system in MATPOWER (Zimmerman et al., 2011), by randomly perturbing the operating conditions (limits for $\mathbf{p}/\mathbf{q}$ and the quadratic cost coefficients of c_i). GNN models with high-order graph filter (Owerko et al., 2020) and `relu` activation for each layer were implemented by PyTorch library. GNN models, and the benchmark FCNN and GiDNN mod-

Method	Metric	118ac	118dc	2383dc
GNN+FR	L_2	5.3e-2	6.4e-2	1-6.8e-2
	Feas.	99.3%	99.4%	99.9%
GNN	L_2	5.2e-2	6.1e-2	6.8e-2
	Feas.	99.1%	99.0%	99.9%
FCNN+FR	L_2	4.5e-2	4.5e-2	4.5e-2
	Feas.	98.9%	90.7%	99.9%
FCNN	L_2	4.4e-2	4.4e-2	4.5e-2
	Feas.	98.8%	90.7%	99.9%
GiDNN+FR	L_2	4.4e-2	5.5e-2	5.6e-2
	Feas.	99.4%	99.4%	99.9%
GiDNN	L_2	5.3e-2	5.0e-2	5.6e-2
	Feas.	99.3%	98.5%	99.9%

Table 1. Performance of GNN, FCNN, and GiDNN in predicting price and solution feasibility.

els (implemented by PyTorch library as well) were tested on Google Colaboratory using the Nvidia Tesla V100 for training acceleration.

Figure 1 compares the performance of proposed GNN-based models with FCNN and GiDNN ones, including those using the feasibility regularized (FR) loss function in (8). The normalized L_2 error in predicting π and the violation rate of line flow limits are considered, for the ac-OPF of 118-bus system and dc-OPF of 2383-bus system. Detailed results are shown in Table 1. Note that for the 2383dc case, although the feasibility results are almost perfect (above 99.9%), learned models with feasibility regularization (FR) in general lead to much smaller infeasibility rates by two orders of magnitude for all three models. Clearly, the performance of proposed GNN models is comparable to that of FCNN and GiDNN ones. The FR loss function design has shown to improve the feasibility of OPF predictions for the larger 2383-bus system. In addition, it can accelerate the training process as corroborated by the actual number of epochs for convergence (not included due to page limit). Compared to FCNN models, GNN ones clearly improve the learning accuracy and feasibility in the 2383-bus system prediction. Hence, the proposed GNN architecture along with feasibility based loss function design has shown effective in predicting feasible OPF solutions, especially for large-scale systems. To demonstrate GNN's reduced complexity, Figure 2 compares the total number of parameters for each model. Detailed comparisons of the total model parameters, occupied memories, training times, and number of training epochs of all models are summarized in Table 2. Although the memory and training time greatly depend on the solvers for different models, the parameter number is indeed reduced by utilizing the topology-based structure of the GNN architecture.

Topology adaptivity: We have further tested the 118-dc

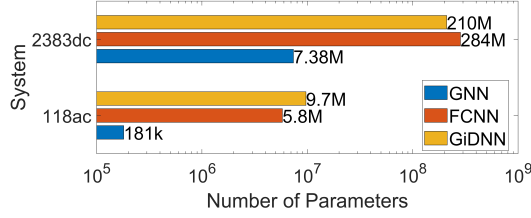


Figure 2. The model complexity of GNN, FCNN, and GiDNN in number of parameters of 118ac and 2383dc systems.

Method	Metric	118ac	118dc	2383dc
GNN+FR	Parameters	351K	351K	142M
	Memory	1.33GB	1.34GB	1.87GB
	Time	29s	95s	103s
	Epoch	49	69	54
GNN	Parameters	351K	351K	142M
	Memory	0.91GB	0.92GB	1.87GB
	Time	35s	160s	159s
	Epoch	114	134	64
FCNN+FR	Parameters	5.8M	5.8M	284M
	Memory	0.91GB	0.93GB	2.24GB
	Time	24s	12s	166s
	Epoch	120	30	35
FCNN	Parameters	5.8M	5.8M	284M
	Memory	0.93GB	0.93GB	2.24GB
	Time	22s	33s	146s
	Epoch	155	90	40
GiDNN+FR	Parameters	9.7M	9.4M	210M
	Memory	1.00GB	1.00GB	2.48GB
	Time	27s	36s	112s
	Epoch	79	94	39
GiDNN	Parameters	9.7M	9.4M	210M
	Memory	1.00GB	1.00GB	2.48GB
	Time	18s	58s	92s
	Epoch	94	104	49

Table 2. Model complexity in the number of parameters and memory, training information in time, and number of epochs of GNN, FCNN, and GiDNN (with or without feasibility regularization) on different systems.

OPF case to validate the topology adaptivity of proposed GNN-based models. Specifically, after obtaining the trained GNN model for the nominal topology, we randomly pick at most two lines to disconnect and test the pre-trained GNN models on this new topology. Figure 3(a) shows that pre-trained GNN models attain satisfactory prediction performance for some new topologies. In addition, we have implemented a post-processing step by using the pre-trained GNNs as a warm start for re-training under each new topology. The post-processing step attains very fast convergence within just 3 – 5 epochs, and high prediction performance as shown in Figure 3(b). This result demonstrates that GNN models are promising in adapting to real-time power grid topology, and points to an exciting direction for future re-

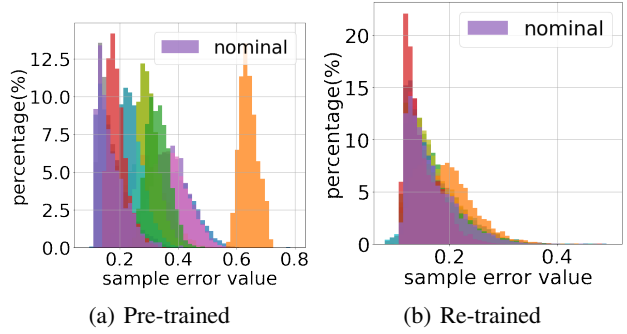


Figure 3. The distribution of sample L_2 prediction error of (a) the pre-trained GNN on randomly perturbed grids and (b) after fast re-training. Each color indicates a new topology.

search .

5. Conclusion and Future Work

This paper proposes a new GNN-based approach for predicting the electricity market prices in order to support the efficient and reliable operations of low-carbon electric grids. Different from earlier learning-for-OPF approaches, our proposed method innovatively incorporates electricity prices' locality property and physics-based regularization term to the design of topology-aware GNN models. Reduced model complexity and topology adaptivity are attained by the GNN-based price prediction. Numerical tests have demonstrated the efficiency and adaptivity of our price prediction method. Interesting future research directions open up on the formal investigation of topology adaptivity and other transfer learning aspects, as well as the extension to general optimal resource allocation problems in networked systems.

Acknowledgements

This work has been supported by NSF Award ECCS-1802319. The authors would like to thank reviewers for providing helpful comments.

References

- Babaeinejadsarookolaei, S., Birchfield, A., Christie, R. D., Coffrin, C., DeMarco, C., Diao, R., Ferris, M., Fliscounakis, S., Greene, S., Huang, R., et al. The power grid library for benchmarking ac optimal power flow algorithms. *arXiv preprint arXiv:1908.02788*, 2019.
- Baker, K. Learning warm-start points for ac optimal power flow. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE, 2019.
- Birchfield, A. B., Xu, T., Gegner, K. M., Shetye, K. S., and Overbye, T. J. Grid structural characteristics as validation

- criteria for synthetic networks. *IEEE Transactions on power systems*, 32(4):3258–3265, 2016.
- Boyd, S., Boyd, S. P., and Vandenberghe, L. *Convex optimization, Ch. 3 & Ch. 5.5*. Cambridge university press, 2004.
- Cain, M. B., O’neill, R. P., Castillo, A., et al. History of optimal power flow and formulations. *Federal Energy Regulatory Commission*, 1:1–36, 2012.
- Castillo, A. and O’Neill, R. P. Survey of approaches to solving the acopf. *Federal Energy Regulatory Commission, Tech. Rep*, 11, 2013.
- Chen, Y. and Zhang, B. Learning to solve network flow problems via neural decoding. *arXiv preprint arXiv:2002.04091*, 2020.
- Deka, D. and Misra, S. Learning for dc-opf: Classifying active sets using neural nets. In *2019 IEEE Milan PowerTech*, pp. 1–6. IEEE, 2019.
- Gama, F., Isufi, E., Leus, G., and Ribeiro, A. Graphs, convolutions, and neural networks: From graph filters to graph neural networks. *IEEE Signal Processing Magazine*, 37(6):128–138, 2020.
- Garcia, M. J. *Non-convex myopic electricity markets: the AC transmission network and interdependent reserve types, Ch. 5 & Ch. 6*. PhD thesis, 2019.
- Garg, V., Jegelka, S., and Jaakkola, T. Generalization and representational limits of graph neural networks. In *International Conference on Machine Learning*, pp. 3419–3430. PMLR, 2020.
- Geng, X. and Xie, L. Learning the lmp-load coupling from data: A support vector machine based approach. *IEEE Transactions on Power Systems*, 32(2):1127–1138, 2016.
- Gholami, A., Ansari, J., Jamei, M., and Kazemi, A. Environmental/economic dispatch incorporating renewable energy sources and plug-in vehicles. *IET Generation, Transmission & Distribution*, 8(12):2183–2198, 2014.
- Guha, N., Wang, Z., Wytock, M., and Majumdar, A. Machine learning for ac optimal power flow. In *n Climate Change Workshop at The Thirty-sixth International Conference on Machine Learning (ICML)*, 2019.
- Isufi, E., Gama, F., and Ribeiro, A. Edgenets: Edge varying graph neural networks. *arXiv preprint arXiv:2001.07620*, 2020.
- Jia, L., Kim, J., Thomas, R. J., and Tong, L. Impact of data quality on real-time locational marginal price. *IEEE Transactions on Power Systems*, 29(2):627–636, 2013.
- Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- Li, M., Wu, Q., Liao, H., Tang, W., and Xue, Y. Optimal power flow with environmental constraints using paired bacterial optimizer. In *2011 IEEE Power and Energy Society General Meeting*, pp. 1–6. IEEE, 2011.
- Ma, Y. and Tang, J. *Deep Learning on Graphs*. Cambridge University Press, 2020.
- Misra, S., Roald, L., and Ng, Y. Learning for constrained optimization: Identifying optimal active constraint sets. *arXiv preprint arXiv:1802.09639*, 2018.
- Owerko, D., Gama, F., and Ribeiro, A. Optimal power flow using graph neural networks. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5930–5934, 2020. doi: 10.1109/ICASSP40776.2020.9053140.
- Pan, X., Zhao, T., and Chen, M. Deepopf: Deep neural network for dc optimal power flow. In *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–6. IEEE, 2019.
- Price, J. E. and Goodin, J. Reduced network modeling of wecc as a market design prototype. In *2011 IEEE Power and Energy Society General Meeting*, pp. 1–6. IEEE, 2011.
- Ramakrishna, R. and Scaglione, A. Grid-graph signal processing (grid-gsp): A graph signal processing framework for the power grid. *IEEE Transactions on Signal Processing*, pp. 1–1, 2021. doi: 10.1109/TSP.2021.3075145.
- Shao, M. and Jewell, W. Co 2 emission-incorporated ac optimal power flow and its primary impacts on power system dispatch and operations. In *IEEE PES General Meeting*, pp. 1–8. IEEE, 2010.
- Wood, A. J., Wollenberg, B. F., and Sheblé, G. B. *Power generation, operation, and control, Sec. 3.10*. John Wiley & Sons, 2013.
- Yang, J., Feng, X., Tang, Y., Yan, J., He, H., and Luo, C. A power system optimal dispatch strategy considering the flow of carbon emissions and large consumers. *Energies*, 8(9):9087–9106, 2015.
- Zamzam, A. S. and Baker, K. Learning optimal solutions for extremely fast ac optimal power flow. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–6. IEEE, 2020.

Zamzam, A. S. and Sidiropoulos, N. D. Physics-aware neural networks for distribution system state estimation. *IEEE Transactions on Power Systems*, 2020.

Zimmerman, R. D., Murillo-Sánchez, C. E., and Thomas, R. J. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1):12–19, 2011.