AP-GNN: Unsupervised Adaptive Distribution Grid-Level Representation Learning

Yuzhou Chen
Department of Computer and Information Sciences
Temple University
yuzhou.chen@temple.edu

Miguel Heleno
Energy Storage and Distributed Resources Division
Lawrence Berkeley National Laboratory
miguelheleno@lbl.gov

Alexandre Moreira
Energy Storage and Distributed Resources Division
Lawrence Berkeley National Laboratory
amoreira@lbl.gov

Yulia R. Gel
Department of Mathematical Sciences
University of Texas at Dallas
National Science Foundation
ygl@utdallas.edu

***Abstract*—Recent episodes of extreme natural events have challenged the ability of power grids to supply demand. Given the increase in the frequency and severity of these events, new methods to evaluate the level of security of power systems are needed. By introducing the emerging deep learning concepts such as attention mechanism for graph neural networks (GNNs) into the power system analysis, we develop a novel approach that systematically classifies this level of security along multiple dimensions. In particular, in contrast to the traditional risk-neutral reliability assessment procedures which focus only the impact of routine failures, our attention-based distribution grid-level representation learning model (AP-GNN) also allows us to simultaneously address the consequences of high impact low probability (HILP) events and to perform unsupervised classification of the distribution grid expansion plans in a computationally efficient manner. Furthermore, we discuss a new tractable resilience metric called Uniqueness Scores which systematically accounts for the key topological characteristics of the heterogeneous distribution grid networks. Our extensive numerical experiments on 54-bus system indicate that the proposed AP-GNN framework is highly competitive both in terms of classification performance and computational efficiency, thereby opening further paths for integration of the state-of-the-art deep learning and artificial intelligence tools to resilience quantification of power systems.**

I. Introduction

Adverse natural events increasingly more often exhibit a negative impact on power grids, challenging their ability to supply demand. The significant rise in the frequency of these events poses the need for new methods that can assess potential expansion plans for distribution systems in terms of security. In this context, we propose a novel method to classify the level of security of potential expansion plans. This method measures the consequences of high impact low probability (HILP) events by computing the Conditional Value at Risk (CVaR) of the system power imbalance, which can take into account the effects of HILP events despite their low probabilities. We present a novel network resilience metric called Uniqueness Scores and a Graph Neural Network (GNN)-based model which can avoid the otherwise necessary solution of multiple computationally demanding optimal power flow (OPF) problems (each one associated with a given scenario of failure) to calculate the CVaR of the system power imbalance corresponding to each potential expansion plan. More specifically, we develop a new attention-based distribution grid-level representation learning model (AP-GNN) for unsupervised classification of distribution grid expansion plans. We also introduce the attention mechanism over the extracted distribution grid-level representations of the distribution grids which allows us to learn adaptive importance weights of the embeddings. We illustrate the proposed method with a 54-bus system and the results corroborate that our AP-GNN approach allows us to improve classification performance, yielding significant gains when compared to the existing state-of-the-art tools.

The key novelty of our contributions are:

- We propose a novel classification algorithm to evaluate the security level of expansion plans for distribution grids, based on a new GNN architecture with a power-based graph convolutional layer and the attention mechanism – Adaptive Power-based Graph Neural Networks (AP-GNN).
- We discuss a new graph-theoretic resilience metric *Uniqueness Score* that accounts for the intrinsic heterogeneity of distribution grid networks.
- In our expansive numerical experiments, we show that AP-GNN yields substantial gains in computational efficiency compared to the stochastic optimization benchmarks from the power system community and also outperforms seven state-of-the-art models from the deep learning community by a significant margin in terms of graph classification accuracy. We then discuss how AP-GNN can be further integrated with power distributional planning tools for operational use.

II. Related Work

Graph Neural Networks for Power System Analysis A wide variety of GNNs have been proposed in recent years for the classification of non-Euclidean structures (see reviews in [1], [2]). However, applications of GNNs in power system analysis remain scarce for both transmission and distribution systems, with only few papers considering GNNs for distribution networks. Such representative approaches include recurrent GCN by [3] for multi-task transient stability assessment of power transmission systems, optimal power flow (OPF) optimization problem in transmission systems [4], node classification in transmission systems [5], and forecasting transmission system responses to contingencies [6]. Taking into account the hidden and complex topological information of distribution systems, some studies [7]–[9] incorporate hypergraph structures and local topological features into GNN-based models for supervised learning tasks. Another application of GNNs to distribution systems include [10] who develop a GCN for fault location in distribution systems, and prove that GCN is more robust to measurement errors compared to the prevailing ML approaches. Compared to these existing techniques, our approach brings multiple new research directions. First, we represent graphical properties of the power distribution grid by local topological signatures and propose a new unsupervised GNN-based model with an attention mechanism. This is a novel application of GNNs not only in power systems but graph learning, in general. Our AP-GNN is inspired by current GNNs but is carefully designed to capture node information with different spatial scopes from a topology distribution grid.

Methods for Distribution Grid Planning Distribution grid planning exercises aim to determine appropriate expansion plans according to the requirements of the system planner, which could be to improve the security of the system or to address future load growth. Distribution system planners usually follow an often empirical process based on codes to update their grids. This process involves electrical installation rules, and a continuous dialog with the regulators to justify the real need for the proposed investments, which can have an impact on the rates paid by consumers [11]. This is usually the case because, in the US, the largest portion of investments made by investor-owned utilities in power systems is associated with distribution grid assets [12].

The power systems research community has proposed a number of relevant methods based on mathematical optimization to improve the distribution grid planning process while considering either reliability or resilience. For example, [13] proposes a linear programming method to determine investments in storage devices with the aim to make the distribution system resilient against earthquakes. In addition, the work developed in [14] has the objective of selecting investments in generation and lines while taking reliability into account. Moreover, [15] considers investments in substations, lines, and transformers within a mixed-integer optimization problem to reduce the cost of loss of load. Finally, [16] presents a mixed-integer stochastic model to optimize investments in distributed generation, line hardening, and line switches so as to reduce costs of expected loss of load and operation of distribution grids. The main advantage of this class of models is its optimal prescriptive nature which allows for a clear justification of the quality of the obtained solutions. However, due to the associated computational burden, it is not straightforward to develop generalizable and standardized industry solutions based on these models for large-scale planning of distribution grids.

Simulation-based methods are often applied to evaluate the quality of potential expansion plans once they are available. Essentially, these methods assess the behavior of the distribution system (or a future system) under the realization of several scenarios (which can be characterized by different failures, load profiles, etc) so as to ensure the technical and economic viability of a given grid planning solution, which could have been obtained either empirically or via an approach based on mathematical optimization. In this context, reliability indices are usually obtained as a result of these simulation methods [17]. For the purpose of evaluating resilience, existing approaches include Monte Carlo methods that comprise a probabilistic characterization of events aided by relevant information about fragility curves, which is used to generate network failures and evaluate their HILP-associated impacts via the CVaR of loss of load [18]. Typically, *every single state of the system* is evaluated in these simulations, considering operational decisions to have the least amount of loss of load each time under several scenarios. Given its nature, this process is highly computationally demanding since it requires the solution of a sequence of multiple optimization problems for large-scale systems.

Hence, in this paper, we propose a computationally efficient method based on GCN that computes the CVaR of loss of load, as a resilience metric, associated with each potential expansion plan. The proposed method, which is intended to replace simulation-based evaluations, is a fast and efficient AI-based evaluation model to assess the resilience of distribution grids that can substantially benefit regulators in their decision making process.

III. METHODS

In this work, our goal is to introduce the power of unsupervised learning of the GNN-based model to classification of distribution grid expansion plans in terms of resilience performance. In particular, our analysis aims at assessing the ability to approximate explicit risk-based resilience metrics, such as the CVaR of load loss, through incorporating a new local graph descriptor into a GNN-based model. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ be a distribution grid, where $\mathcal{V}$ is a set of nodes ($|\mathcal{V}| = \mathcal{N}$), $\mathcal{E}$ is a set of edges, $A \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ is a symmetric adjacency matrix and $X \in \mathbb{R}^{\mathcal{N}\times F}$ is a node feature matrix (where F is the dimension of node features). In a distribution grid, each node represents either a bus or a substation, and each edge is a distribution line between nodes. The weight of an edge between nodes u and v in $\mathcal{G}$ is denoted by e_{uv}, with $e_{uv} = 0$ if there exist no path connecting nodes u and v.

A. Base Resilience Evaluation Method

Methods that assess reliability and resilience are designed to evaluate the capability of the power grid to withstand multiple scenarios of outages while effectively supplying load. To account for pre-defined sets of scenarios of routine failures and HILP events, the resilience evaluation computes the distribution of total system loss of load (i.e., the sum of loss of load across all nodes of the system) and then calculates the CVaR of this distribution. This process is usually carried out via Sequential Monte Carlo Simulation (SMCS) methods, which perform simulations of loss of load over annual time horizons. More specifically, to compute the annual CVaR of loss of load, we simulate the operation of the distribution system for several scenarios (e.g., 2000 scenarios), where each scenario corresponds to 365 days and the system is operated for each hour of each day. Due to its respective failures, each annual scenario is then associated with a particular annual loss of load. Given the amounts of loss of loads associated with all the scenarios, we then calculate the annual CVaR of loss of load. This procedure can take 24 hours for a 54-Bus system to evaluate the annual CVaR for a *single candidate expansion plan*. In practice, the power grid expansion plans are noticeably larger, and hence, the associated computational costs in operational settings are substantially higher. Hence, one of the primary contributions of this paper is *to replace the computationally expensive simulation by a neural networks framework that can efficiently classify distribution grids according to their corresponding ranges of annual CVaR of loss of load*.

B. Uniqueness Scores

Traditional resilience metrics used for analyzing power grid networks are based on graph theory, such as average clustering coefficient, average path length, average betweenness centrality, diameter, and transitivity [19], [20]. Nevertheless, existing metrics used in power grid analysis suffer from limitations as they lack the ability to distinguish between different types of nodes, such as substations and buses, and primarily emphasize global network characteristics while neglecting crucial local graph topology information. To address this gap, our study aims to introduce a novel node feature for load analysis based on an in-depth examination of simple paths connecting loads and substations. Drawing inspiration from the concept of average path length, we utilize a new network summary to achieve this goal, i.e., **Uniqueness Scores (U-scores)** [7] to capture the efficiency of information transport between different loads and substations. The U-scores algorithm is summarized in a framework for distribution system expansion planning, as depicted in Algorithm 1. Specifically, the U-scores framework consists of three main components: (1) extracting simple paths from one source, (2) calculating scores, and (3) finalizing uniqueness scores. Each component is discussed in detail below. Given the original graph structure $\mathcal{G}_{\text{Bus}}$ and the expansion plan $\mathcal{G}_{\text{Exp}}$, $\mathcal{G}_{\text{Comb}}$ represents the combination of G_{Bus} and $\mathcal{G}_{\text{Exp}}$. The algorithm for U-scores calculation is shown in Algorithm 1.

Furthermore, in order to effectively assess the robustness of the distribution system, the introduced U-scores metric provides invaluable insights into *how each load impacts the resilience of the complex network of the distribution system*. This quantitative evaluation offers several advantages. Firstly,

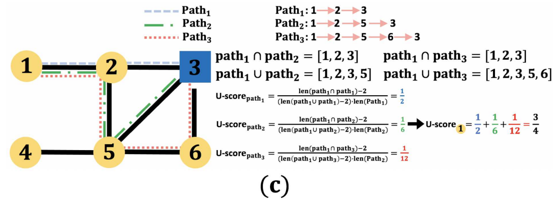

Fig. 1. A toy example of the U-scores computation for load 1 on substation 3.

it enhances the understanding of load dynamics for power system operators, enabling them to develop more efficient and proactive risk mitigation strategies. Secondly, the U-scores metric enables electrical engineers to accurately and reliably evaluate potential expansion plans for existing distribution systems, by comparing the summation of U-scores (i.e., higher U-scores summation indicates more resilient distribution systems). Figure 1 illustrates a toy example of the computation of U-scores for load 1 on substation 3 in a distribution grid.

Algorithm 1: Uniqueness Scores (U-scores)

Input: Bus system G_{Bus} and expansion plan G_{Exp}
Output: Uniqueness scores $\mathbb{S} \in \mathbb{R}^{\mathcal{N}_{\text{load}} \times \mathcal{N}_{\text{substation}}}$

```
G_Comb = G_Bus ∪ G_Exp
for i ← 1 to N_substation do
    for j ← 1 to N_load do
        simplePaths_{l_j,s_i} ← DFS(G_Comb, l_j, s_i);
        simplePaths^sorted_{l_j,s_i} ← sort(simplePaths_{l_j,s_i});
        S⃗_{l_j,s_i} ← (); U_{l_j,s_i} ← {}; I_{l_j,s_i} ← ∅;
          P_{l_j,s_i} ← 0;
        for k ← 1 to len(simplePaths^sorted_{l_j,s_i}) do
            if k == 1 then
                S^(1)_{l_j,s_i} = 1;
                U_{l_j,s_i} = U_{l_j,s_i} ∪ simplePaths^sorted_{l_j,s_i}[k];
                P_{l_j,s_i} =
                  P_{l_j,s_i} + len(simplePaths^sorted_{l_j,s_i}[k]);
            end
            else
                I_{l_j,s_i} = simplePaths^sorted_{l_j,s_i}[k] ∩ U_{l_j,s_i};
                U_{l_j,s_i} = U_{l_j,s_i} ∪ simplePaths^sorted_{l_j,s_i}[k];
                S^(k)_{l_j,s_i} = 1 − (len(I_{l_j,s_i}) − 2) / (len(simplePaths^sorted_{l_j,s_i}[k]) − 2);
                P_{l_j,s_i} =
                  P_{l_j,s_i} + len(simplePaths^sorted_{l_j,s_i}[k]);
            end
            S[j, i] = sum(S⃗_{l_j,s_i}, axis=0) / P_{l_j,s_i};
        end
    end
end
```

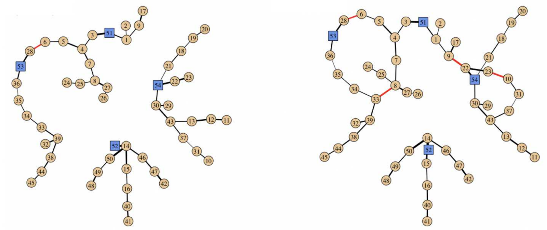

Fig. 2. Expanded systems with two expansion plans, where node the represents the load, □ node represents the substation, the black edge represents the existing line, and red edge represents the added candidate line.

C. Graph Neural Networks with a Power-based Layer and Attention Mechanism

Graph-based semi-supervised learning (G-SSL) is arguably the most popular approach for graph-based supervised learning in recent years. G-SSL develops a generalized optimization framework, which has three particular cases (i) the Standard Laplacian (SL); (ii) the Normalized Laplacian (NL); (iii) PageRank (PR). Optimization formulation with the following expression is to find an $N \times K$-matrix Z (i.e., prediction matrix, or classification function; where K is the number of classes):

$$\min_{Z} \left\{ 2Z^T D^{\sigma-1} L D^{\sigma-1} Z + \mu (Z-Y)^T D^{2\sigma-1}(Z-Y) \right\},$$

where μ is a regularization parameter. Minimization of the 1st term in the expression above corresponds to Laplacian regularization; and by minimizing the 2nd term we aim to use the least-square-fitting method. In turn, the optimization formulation above allows us to obtain (i) the Standard Laplacian based formulation ($\sigma = 1$), (ii) the Normalized Laplacian formulation ($\sigma = 1/2$), and (iii) PageRank formulation ($\sigma = 0$). The objective of the generalized optimization framework for G-SSL is a convex function and the corresponding the classification function, i.e., $Z = (1-\alpha)\left(I - \alpha D^{-\sigma} W D^{\sigma-1}\right)^{-1} Y$, where $\alpha = 2/(2+\mu)$. From the above formulations, classification function Z is a closed form solution based on the theory of random walks on graphs, which in turn provides connection to the probabilistic interpretation of G-SSL. Parameter α controls the strength of the ground truth label matrix Y in the generalized optimization framework.

Adaptive Power-based Graph Neural Networks Framework Our proposed network is Adaptive Power-based Graph Neural Networks (AP-GNN), because the power-based graph convolution layer (PGCLayer) is able to incorporate the power information from substations to load node representation and the attention mechanism can adaptively learn the deep correlation information between different node embeddings. Algorithm 2 represents the operations in PGCLayer. After ℓ-th iteration, the node representation in graph $\mathcal{G}_i$ can be defined as (line 4 of Algorithm 2)

$$h^{(\ell-1)}_{i,\mathcal{N}_k(v)} = \text{AGG}\left(h^{(\ell-1)}_{i,u}, \forall u \in \mathcal{N}_k(v)\right),$$

where AGG$(\cdot)$ is the aggregation function that aggregates the output of each neighbor (e.g., sum and $mean$), $\mathcal{N}_k(v)$ refers to k-hop neighbors of node v including v itself (where $k \geq 1$), and the neighborhood vector $h^{(\ell-1)}_{i,\mathcal{N}_k(v)}$ incorporates node feature information from v's neighborhood into the representation. Note that k-th power contains the spatial scope from the k-th step of a random walk on the target distribution grid. To capture the dependencies between load and substation, PGCLayer also can gather information from substations via

$$h^{(\ell-1)}_{i,\text{substations}} = \text{AGG}\left(h^{(\ell-1)}_{i,s}, \forall s \in \mathcal{N}_{\text{substation}}\right),$$

that is, line 5 of Algorithm 2, i.e., for load node, we learn the node embedding based on information from both 1-hop neighborhood and all substations (line 7 of Algorithm 2). The core ideas behind the new graph representation are to (i) utilize U-scores information (which provides the quantitative measures between the load and all substations) and (ii) assign more weights to nodes that have connections with substations. To explore neighborhoods of nodes at different depths, we consider K different steps of random walks (i.e., $1 \leq k \leq K$). Therefore, we can utilize PGCLayer to extract K node embeddings from the target distribution grid $\mathcal{G}_i$, i.e., $\{H_1, \ldots, H_k, \ldots, H_K\}$ where H_k represents the node embedding of PGCLayer with the k-th step of a random walk (for the sake of simplicity, we omit ℓ that represents a layer in the following discussion).

We employ the attention mechanism to dynamically capture the intrinsic relationships between various node embeddings, allowing us to focus on the crucial task-relevant components of the learned representations, thereby facilitating informed decision-making. More specifically, we define the attention mechanism as follows $(\alpha_{H_1}, \alpha_{H_2}, \cdots, \alpha_{H_K}) = Attention(Z_{H_1}, Z_{H_2}, \cdots, Z_{H_K})$, and $\alpha_{H_k} = \text{softmax}_i(\varrho_{\text{Att}} \tanh(\Theta H_k))$, where ϱ_{Att} is a linear transformation, Θ is the trainable weight matrix, and the softmax function is used to normalize the attention vector. We then can obtain the final embedding H by combining all embeddings with the attention weights

$$H = \alpha_{H_1} \times H_1 + \alpha_{H_2} \times H_2 + \cdots + \alpha_{H_K} \times H_K.$$

Finally, the final embedding H given by the AP-GNN is fed into a multilayer perceptron (MLP) layer. Moreover, inspired by [21], we use mutual information maximization for unsupervised representation learning on the target distribution grid.

IV. Experimental Studies

A. Datasets

We consider learning CVaR of Annual Loss of Load through multi-class classification. Each of these classes represents a "degree" of system risk in relation to HILP events. The resulting classification could be then converted into regulatory planning standards, allowing to systematically compare resilience between different plans of the same grid and potentially across grids in similar geographical conditions [27]. Following a standard statistical practice [28], we perform binning into classes based on quantile ranges of annual CVaR of loss of load (kWh) (see Table II). In particular, we get 3 classes

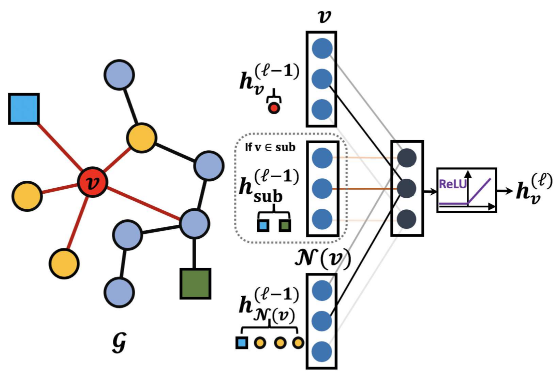

Fig. 3. Overview of a power-based graph convolution layer. The red node v is the target node, yellow nodes represent the 1-hop neighborhood of v, and ☐ nodes represent substation. Traditional graph convolution learns the node representation via the node self and its neighbors, while our method can build connections between specific nodes (i.e., load and substation) and avoid expansive neighborhood expansion.

Algorithm 2: Power-based Graph Convolution Layer

Input: Data $\boldsymbol{X} = \{X_i\}$ and $\boldsymbol{\mathcal{G}} = \{\mathcal{G}_i\}$

for i-th graph $\{\mathcal{G}_i, X_i\}$ in mini-batch **do**
- $h_i^{(0)} = f_{\text{MLP}}(X_i)$
- **for** $\ell \leftarrow 1$ **to** nlayers **do**
  - $h_{i,\mathcal{N}_k(v)}^{(\ell-1)} = \text{AGG}\left(h_{i,u}^{(\ell-1)}, \forall u \in \mathcal{N}_k(v)\right)$
  - $h_{i,\text{substations}}^{(\ell-1)} = \text{AGG}\left(h_{i,s}^{(\ell-1)}, \forall s \in \mathcal{N}_{\text{substation}}\right)$
  - **if** $v \notin \mathcal{N}_{\text{substation}}$ **then**
    - $h_{i,v}^{(\ell)} = \sigma\left(W \cdot \bigoplus\left(h_{i,\mathcal{N}(v)}^{(\ell-1)}, h_{i,\text{substation}}^{(\ell-1)}\right)\right)$
  - **end**
  - **else**
    - $h_{i,v}^{(\ell)} = \sigma\left(W \cdot h_{i,\mathcal{N}_k(v)}^{(\ell-1)}\right)$
  - **end**
- **end**

end

TABLE I
AVERAGE ACCURACY (%) COMPARISON WITH BASELINE METHODS.

Model	3 classes		4 classes	
	54-Bus System I	54-Bus System II	54-Bus System I	54-Bus System II
RF [22]	52.99	44.82	51.49	42.68
WL subtree [23]	53.00	43.39	42.25	44.82
GCN [24]	66.50	61.43	71.50	49.46
GIN [25]	76.00	63.75	65.49	53.57
KCNN [26]	75.50	63.29	71.30	55.00
AP-GNN (ours)	**77.49**	**63.93**	**72.50**	**56.43**

(i.e., low-, moderate-, and high-risk) and 4 classes (i.e., low-, moderate-, middle-, and high-risk). We have implemented our proposed methodology on two distribution systems, i.e., 54-Bus System I and 54-Bus System II, which are modified versions of the 54-Bus system as described in [14]. In the 54-Bus System I, there are a total of 72 lines, comprising of 50 existing lines and 22 candidate lines. Additionally, the system consists of 4 substation nodes and 50 load nodes. On the other hand, the 54-Bus System II includes 72 lines, with 52 existing lines and 20 candidate lines. Furthermore, the system contains 2 substation nodes, 50 load nodes, and 2 non-load nodes. To comprehensively evaluate the performance of our proposed AP-GNN model, we generate multiple expansion plans for these two systems, resulting in 200 networks for the 54-Bus System I and 74 networks for the 54-Bus System II. These expansion plans have been generated by selecting different subsets of the available candidate lines. For more details of datasets, please refer to [7].

TABLE II
RANGES OF ANNUAL CVAR OF LOSS OF LOAD.

# Classes	Label	54-Bus System I Range (kWh)	54-Bus System II Range(kWh)
# 3 classes	0: low-risk	[0, 1.0e4]	[0, 3.0e4]
	1: moderate-risk	(1.0e4, 2.0e4]	(3.0e4, 4.0e4]
	2: high-risk	(2.0e4, ∞)	(4.0e4, ∞]
# 4 classes	0: low-risk	[0, 1.0e4]	[0, 3.0e4]
	1: moderate-risk	(1.0e4, 2.0e4]	(3.0e4, 3.5e4]
	2: middle-risk	(2.0e4, 3.0e4]	(3.5e4, 4.0e4]
	3: high-risk	(3.0e4, ∞]	(4.0e4, ∞)

B. Experimental Settings and Baselines

We use the Adam optimizer for 100 epochs to train AP-GNN. For both 54-Bus Systems, AP-GNN consists of 5 layers whose hidden feature dimension which varies with the range $\{8, 16, 32, 64, 128\}$, and each layer consists of two MLP blocks. The learning rate is 0.01 the dropout is set as 0.5, and the batch size is set as 16. For k-th step of a random walk, we set the largest step of a random walk $K \in \{2, 3, 4\}$. The best results are in **bold** font. We implement our proposed AP-GNN model using Pytorch on NVIDIA GeForceX 3090. We compare our AP-GNN model with five state-of-the-art baselines, including (i) Random Forest [22] (RF), (ii) Weisfeiler-Lehman subtree kernel [23] (WL subtree kernel), (iii) Graph Convolutional Networks [24] (GCN), (iv) GIN [25], and (v) Kernel Graph Convolutional Neural Networks [26] (KCNN).

C. Overall Results

Table I shows the comparison of our proposed AP-GNN and state-of-the-art baselines for unsuerpvised graph classification tasks. From Table I, we observe that our AP-GNN consistently outperforms baselines on both 54-Bus System I and 54-Bus System II. Moreover, our proposed AP-GNN outperforms the runner-ups (GIN and GCN) with relative gains of 1.92% and 1.38% on 54-Bus System I with 3 classes and 4 classes respectively. For 54-Bus System II, AP-GNN outperforms the runner-ups (GIN and KCNN) with relative gains of 0.28% and 2.53% over 3 classes and 4 classes scenarios respectively.

V. Conclusion

Despite their high proliferation into a broad range of disciplines from social networks to bioinformatics to finance, GNNs and more generally DL tools still remain largely underutilized in the context of analysis of electricity distribution grids and distribution grid resilience planning, in particular. In this paper we have introduced such emerging DL concepts as attention mechanisms for graph learning into the unsupervised classification of the future distribution grid expansion plans in terms of their response to routine and HILP events. Our numerical experiments have shown that the proposed GNN tool with the attention-based mechanism delivers highly competitive results both in terms of classification performance and computational costs. These results suggest that the proposed AP-GNN and similar DL methods for graph learning may constitute a new promising alternative for automatic classification of future distribution grids as a pre-solver for the computationally expensive stochastic expansion and planning optimization methods.

Acknowledgment

This work was authored in part by the Lawrence Berkeley National Laboratory (LBL), U.S. Department of Energy (DOE) National Laboratory operated by the University of California and also was supported by the NSF grant # ECCS 2039701, INTERN supplement for ECCS 1824716, ONR grant # N00014-21-1-2530, and DOE # DE-AC02-05CH11231. The views expressed in the article do not necessarily represent the views of the DOE, NSF, and ONR.

References

[1] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.

[2] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE TKDE*, 2020.

[3] J. Huang, L. Guan, Y. Su, H. Yao, M. Guo, and Z. Zhong, "Recurrent graph convolutional network-based multi-task transient stability assessment framework in power system," *IEEE Access*, vol. 8, pp. 93 283–93 296, 2020.

[4] D. Owerko, F. Gama, and A. Ribeiro, "Optimal power flow using graph neural networks," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2020, pp. 5930–5934.

[5] Y. Chen, Y. Gel, and K. Avrachenkov, "LFGCN: Levitating over graphs with levy flights," in *IEEE International Conference on Data Mining*. IEEE, 2020, pp. 960–965.

[6] B. Bush, Y. Chen, D. Ofori-Boateng, and Y. Gel, "Topological machine learning methods for power system responses to contingencies," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 15 278–15 285.

[7] Y. Chen, T. Jiang, M. Heleno, A. Moreira, and Y. R. Gel, "Evaluating distribution system reliability with hyperstructures graph convolutional nets," in *IEEE International Conference on Big Data*, 2022, pp. 1793–1800.

[8] Y. Chen, M. Heleno, A. Moreira, and Y. R. Gel, "Topological graph convolutional networks solutions for power distribution grid planning," in *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2023.

[9] Y. Chen, R. A. Jacob, Y. R. Gel, J. Zhang, and H. V. Poor, "Learning power grid outages with higher-order topological neural networks," *IEEE Transactions on Power Systems*, pp. 1–13, 2023.

[10] K. Chen, J. Hu, Y. Zhang, Z. Yu, and J. He, "Fault location in power distribution systems via deep graph convolutional networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 119–131, 2019.

[11] A. Cooke, J. Homer, and L. Schwartz, "Distribution system planning - state examples by topic," May 2018, accessed: 26-4-2022. [Online]. Available: https://www.osti.gov/biblio/1561123

[12] EEI, "Industry capital expenditures." Oct. 2019, accessed: 26-4-2022. [Online]. Available: https://eei.org/issuesandpolicy/Finance

[13] M. Nazemi, M. Moeini-Aghtaie, M. Fotuhi-Firuzabad, and P. Dehghanian, "Energy storage planning for enhanced resilience of power distribution networks against earthquakes," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 795–806, 2019.

[14] G. Muñoz-Delgado, J. Contreras, and J. Arroyo, "Multistage generation and network expansion planning in distribution systems considering uncertainty and reliability," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3715–3728, 2016.

[15] G. Muñoz-Delgado, J. Contreras, and J. M. Arroyo, "Distribution network expansion planning with an explicit formulation for reliability assessment," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2583–2596, 2018.

[16] S. Ma, S. Li, Z. Wang, and F. Qiu, "Resilience-oriented design of distribution systems," *IEEE Transactions on Power Systems*, pp. 2880–2891, 2019.

[17] R. Billinton and J. Billinton, "Distribution system reliability indices," *IEEE Transactions on power Delivery*, pp. 561–568, 1989.

[18] P. Gautam, P. Piya, and R. Karki, "Resilience assessment of distribution systems integrated with distributed energy resources," *IEEE Transactions on Sustainable Energy*, pp. 1–1, 2020.

[19] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[20] A. K. Dey, Y. R. Gel, and H. V. Poor, "What network motifs tell us about resilience and reliability of complex networks," *Proceedings of the National Academy of Sciences*, vol. 116, no. 39, pp. 19 368–19 373, 2019.

[21] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, "Learning deep representations by mutual information estimation and maximization," in *Proceedings of the International Conference on Learning Representations*, 2018.

[22] A. Liaw and M. Wiener, "Classification and regression by randomforest," *RNews*, vol. 2, no. 3, pp. 18–22, 2002.

[23] N. Shervashidze, P. Schweitzer, E. J. Van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels." *Journal of Machine Learning Research*, vol. 12, no. 9, 2011.

[24] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of the International Conference on Learning Representations*, 2017.

[25] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *Proceedings of the International Conference on Learning Representations*, 2018.

[26] G. Nikolentzos, P. Meladianos, A. J.-P. Tixier, K. Skianis, and M. Vazirgiannis, "Kernel graph convolutional neural networks," in *International conference on artificial neural networks*. Springer, 2018, pp. 22–32.

[27] J. Ma, J. Zhang, L. Xiao, K. Chen, and J. Wu, "Classification of power quality disturbances via deep learning," *IETE Technical Review*, vol. 34, pp. 408–415, 2017.

[28] F. Nargesian, H. Samulowitz, U. Khurana, E. B. Khalil, and D. S. Turaga, "Learning feature engineering for classification." in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2017, pp. 2529–2535.