

Learning Power Grid Outages with Higher-Order Topological Neural Networks

Yuzhou Chen, Roshni Anna Jacob, *Graduate Student Member, IEEE*, Yulia R. Gel,
Jie Zhang, *Senior Member, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

Abstract—With the increase in cyber-physical threats and extreme weather events, resilience of the power system has become a problem of utmost societal importance. In this paper, we propose a novel approach for resilience improvement of power distribution networks, based on the notions of persistent homology and simplicial neural networks (SNNs) which are new directions in graph learning. In particular, tools of persistent homology allow us to capture the most essential topological descriptors of the distribution network. In turn, extending the convolutional operation to simplicial complexes on the distribution network, using the Hodge-Laplacian analytics, enables us to describe complex interactions among multi-node higher order graph substructures. Such higher order graph substructures are of particular importance in distribution networks, since a change in power demand at a load bus (or the power supplied from a substation) will produce a corresponding perturbation in nodal variables (such as the bus voltages) and edge variables (such as branch currents). We validate our new Higher-Order Topological Neural Networks (HOT-Nets) model for contingency classification of three test distribution networks, the IEEE 37-bus feeder, IEEE 123-bus feeder, and the 342-bus low voltage network. Our experiment results on two case studies (i.e., (i) with sensors placed at all the buses alone in the networks and (ii) with partial observability in the networks) indicate that HOT-Nets substantially outperforms 9 state-of-the-art methods, yielding relative gains of up to 14.04% in terms of system resilience classification.

Index Terms—Resilience, distribution network, graph neural networks, graph learning.

I. INTRODUCTION

A resilient power network is crucial for the security and prosperity of a society. The alarming increase of outages in the power network due to extreme weather events and cyber-physical attacks has further necessitated enhancing the resilience of the power network infrastructure. This includes strategies to prevent, detect, and mitigate outages caused by contingency events. Outage detection in the distribution network (DN) is hence an important facet for improving the

Yuzhou Chen is with the Department of Computer and Information Sciences at Temple University, Philadelphia, PA 19122, USA (e-mail: yuzhou.chen@temple.edu).

Roshni Anna Jacob is with the Department of Electrical and Computer Engineering at The University of Texas at Dallas, Richardson, TX 75080, USA (e-mail: roshni.jacob@utdallas.edu).

Yulia R. Gel is with the Department of Mathematical Sciences at The University of Texas at Dallas, Richardson, TX 75080, USA; and National Science Foundation (e-mail: ygl@utdallas.edu).

Jie Zhang is with the Department of Mechanical Engineering and (Affiliated) Department of Electrical and Computer Engineering at The University of Texas at Dallas, Richardson, TX 75080, USA (e-mail: jiezhang@utdallas.edu).

H. Vincent Poor is with the Department of Electrical and Computer Engineering at Princeton University, Princeton, NJ 08544, USA (e-mail: poor@princeton.edu).

resilience and provides the much needed real-time situational awareness prior to implementing restorative actions [1], such as distributed energy resource dispatch, network reconfiguration [2] and sectionalizing into microgrids [3]. Additionally, the knowledge of the operating status of the DN is critical for state estimation, monitoring of distributed energy resources, and management of demand response [4], [5].

Line failures or short circuit faults in the DN may be caused due to natural reasons (such as aging and extreme weather) or cyber-physical attacks. During such adverse conditions, the protective devices function to automatically isolate the faulted components, and disconnect the downstream network with its connected loads from the main grid, thereby resulting in an outage. Fault diagnostics (determining the location and the type of failure) are often performed after detecting an outage [6]. Hence, outage detection is preeminent and corresponds to the task of finding the status of the network functionality including that of the protective devices [7].

The transmission network is known to be a critical infrastructure and hence has a widespread presence of advanced communication, metering, and sensor equipment (such as phasor measurement units, PMUs). Several techniques have been proposed for outage detection in transmission networks, such as Bayesian regression [8] and sparse vector estimation [9]. In contrast, the DN outage detection is more challenging considering the unobservability of the network [10]. This is because until recently, the DN was overlooked in terms of its criticality which is now changing with the transition towards the smart grid concept and decentralization of the grid. Unlike the transmission network, the DN traditionally has a radial structure, larger reactance-resistance ratios, and is spread out with hundreds of thousands of nodes. Hence, many of the algorithms developed for the transmission networks may not be suitable for DN outage detection. Additionally, increasing the number of sensors to impart global observability to the DN would result in high installation and maintenance costs [11]. On the other hand, with innovations in instrumentation technology, a decrease in total sensor cost is expected over the years. Despite this, considering factors such as latency and communication bottleneck with the DN size [5], limited sensors placed optimally at select locations are deemed to be sufficient to capture the system state. This is because any information relayed to the distribution system operator is desired to be real-time so as to enable instantaneous control and transfer to the emergency management system if necessary.

Sparse sensor placement is an area with a plethora of

research surrounding the advancements in compressed sensing. Sparse sensors and related network classification or signal reconstruction are attributed to the behavior of complex high dimensional dynamic systems to exhibit low dimensional patterns [12]. These patterns can be exploited to learn the system state using data-driven techniques. There exist different streams of data in the DN. The supervisory control and data acquisition (SCADA) system provides the electrical measurements and status of the devices in the primary substations, and is expected to cover the lower voltage networks in the imminent future. Power consumption and generation of customers in the DN are measured using the advanced metering infrastructure (AMI). The smart meters in this class can also provide the voltage and current magnitudes at several points in the network. Others such as micro-PMUs and linewatch sensors provide the real and reactive power flows along the lines incident to the node at which these are placed [13]. Time series measurements from PMUs specifically designed for DNs are also being employed for analytics and diagnostics. Optimal placement of these devices in the DN has already been studied in several works such as [1], [7] and is beyond the scope of our paper.

The data from the sensors can be used for three different categories of tasks, i.e., predictive, prescriptive, and descriptive. In predictive approaches, historical data will be used to predict the state of the network or events in the near future. State estimation is a common task in this category. Switching (network reconfiguration) and load shedding, etc., fall in the class of prescriptive tasks where suitable actions and/or probable outcomes are prescribed based on the current system state. Descriptive approaches are used to learn from past behaviors to classify elements into groups [14]. Fault diagnostics and outage detection are typical examples in this category and usually encompass classification and clustering techniques. In this paper, the authors propose a new graph learning architecture for descriptive tasks, in particular outage detection using available sensor measurements.

Until recently, the DN operators gathered information about the network functionality from customer calls and crew inspections, or using the signals picked up by sensors in the substation [7]. Following this, knowledge based systems utilizing both customer calls and polling from AMI resources were used for identifying DN outages [15]. This however, lacked in a real-time identification of the overall system performance considering the expense of acquiring data and the computational constraints associated with size. Further with the deployment of smart sensors, direct monitoring of the status of some of the network parameters and the substation electrical parameters was made possible. Considering the requirement for timely detection of network contingencies given sparsity in sensors, machine learning (ML) models are considered as a good fit and hence employed in several works such as that in [6].

Extreme events which result in network outages are known to be high-impact, low-probability events. A major limitation associated with this is the non availability of sufficient data (from sensors during outages) to train the ML models. This is often mitigated by the adoption of pseudo measurements (e.g., historical load forecasts and weather predictions) and physics

based (power flow) models. On the other hand, different outage scenarios may have nearly similar observations in power flows, thus resulting in a non-unique mapping, which is challenging to learn [5]. Hence, using a naive ML approach with pseudo measurements and power flow measurements may not be sufficient. The actual power flow is essentially a function of the network topology, load demand, and the outage scenario in the network [7]. This implies that factoring in the topological aspect of the DN is pertinent for the outage detection task and has received relatively little attention in the literature.

The DN has an inherent graph structure, with buses as nodes, and lines or transformers as edges. There also exists a strong interdependence among state variables such as bus voltages, and line flows which can be modeled as node and edge variables respectively. Considering that component failures result in changes in the topology of the distribution feeder, and the underlying network connectivity also affects the scale of system degradation, we adopt a learning-over-graph approach to detect power disruption in the network.

To address the above mentioned limitations, and in particular, to account more accurately for the role of local topological information in the DN as well as the latent multi-node interactions among loads, substation buses and the subsequent branch flows, we propose a novel deep neural architecture, namely, Higher-Order Topological Neural Networks (HOT-Nets) which fuses two emerging directions in graph learning, i.e., a fully trainable topological layer and a convolution operation on simplicial complexes. We extensively validate the HOT-Nets architecture for outage detection on power distribution networks. Our findings show that HOT-Nets delivers significant improvements in the classification of distribution system outages, with relative gains up to 14.04%, compared to 9 state-of-the-art deep learning methods.

The main contributions of this work can be summarized as follows:

- We introduce the concepts of topological signatures and higher order interactions among multi-node graph substructures into learning of power distribution systems, by aggregating the benefits of persistent homology, Hodge theory, and deep neural networks. To the best of our knowledge, neither tools of persistence homology nor Hodge-Laplacians alongside graph neural networks have previously been applied in conjunction with any resilience related task in distribution grids.
- To the best of our knowledge, this is the first deep neural architecture that leverages both the utility of a simplicial convolutional layer and a fully trainable topological layer. Considering the slow adoption of graph learning techniques in power system applications, our work goes even beyond this to leverage tools of topological data analysis in characterizing distribution networks.
- To enable the resilient operation of the distribution network, we develop a model for detecting outages in a practical distribution network with sparse sensors. The model is validated on two conventional radial and one urban, low voltage meshed distribution networks. Additionally, two case studies, i.e., one with sensors placed at all the buses alone in the networks and another one with

partial observability in the networks are performed. The performance of the model is compared with state-of-the-art learning models.

II. RELATED WORK

Outage detection and resilience in power distribution

networks The degradation of the system state with the evolution of high impact, low probability extreme events in the DN is a temporal process, which can be categorized as before, during, and after an event. Hence, the tasks associated with outages can be classified as prediction, detection, and mitigation accordingly. It is important to infer the system status during or immediately following line outages prior to taking restorative actions such as network reconfiguration, intentional islanding [16], etc. Generally, outage detection relies on using sensors allocated in the DN as, for example, discussed in [5], [6].

In related tasks such as state estimation and topology identification, several works have employed classical approaches such as weighted least square and weighted least absolute value methods. In these methods, state estimation or topology identification is modeled as optimization (usually mixed-integer linear programming) problems that can be solved using suitable solvers. In [17], for example, a reformulated version of the weighted least absolute value (WLAV) method has been used for topology identification. Although the WLAV is more robust against bad data (compared to weighted least square optimization methods), they are computationally expensive and sensitive to measurement uncertainty, making it unsuitable for real-time online state estimation. Additionally, the robustness and bad data rejection by the WLAV method are also attributed to measurement redundancy which may not be present in the distribution networks [17]. The distribution networks are often only partially observable due to limited sensor placement in the network. This is because the distribution network is considerably large with hundreds of thousands of nodes/lines and it may not be economically feasible to allocate sensors for each component. However, the limited sensors are placed optimally such that the network state can be derived from the sparse signals. This principle of compressed or sparse sensing is used in [18], however, with a mixed integer non-linear programming approach which is computationally intensive, given the NP-hard nature of the state estimation problem. In contrast, some works use the principle of compressed or sparse sensing as in [18], however, with a mixed integer non-linear programming approach which is computationally intensive, given the NP-hard nature of the state estimation problem.

Considering outage-specific tasks in DN, in [5], a strategy of partitioning DN into sub networks which are further divided into multiple control areas is adopted. Similarly, in [6], optimally placed power flow sensors and the load estimates acquired from the advanced metering infrastructure (AMI) are used for outage detection by dividing the tree network into subtrees with sensors placed at the root and boundaries of subtrees. The detector is based on a maximum a-posteriori probability (MAP) formulation and particularly, the graph (tree) structure of the DN is also considered with the power flow measurements as the edge variables, and the

load forecasts as the node variables. However, as mentioned earlier, taking into account the partial observability of the DN with limited sensors, the non-scalability, and the latency associated with such approaches, ML based techniques can alternatively be adopted for real-time outage detection and situational awareness. In this light, [19], [20] have adopted ML approaches where the historical outage data for specific extreme weather events is collected, analyzed, and used for training a deep neural network to predict repair time and restoration time. Such models are useful for post-disaster decision making such as optimal repair and crew scheduling. Many outage prediction models using weather data and ML models are also available in the literature such as [21], [22]. In particular, a graph embedding model in [22] has been adopted, where a DN overlaying a geographical area has been modeled as a graph to account for the spatial information and the variation in weather with geography, while performing outage prediction. These models can be used for early warning and preparation for outages such as increasing storage, prosumers trading renewable energy [23], etc. Alternatively, the outage detection problem is addressed in [24], where social media (tweets) are used as sensors. Although being a practical and interesting approach, this lacks spontaneous detection as the human response to such events is much slower compared to the speed of outages and cascading effects in DN. There exist models such as [25], where a data-driven scheme with smart meter data and ML models are used for outage detection. Alternatively, in [26], the authors propose a deep learning (DL)-based approach for topology identification (TI) and state estimation for an unbalanced three-phase distribution system. These models are however purely data-driven and do not consider any underlying topological functionalities to accurately capture network dynamics.

As opposed to the large amount of work concerning outage prediction, outage location, and prediction of repair and restoration time, outage detection using ML is comparatively limited. It is, however, essential to know if there are indeed any load outages (i.e., loads not served) and the risks associated with line outages in real time. To enhance the accuracy of the model used for situational awareness, incorporating the topological information along with the physics-based power flow information for each outage scenario in the learning framework has not yet been explored. This is necessary especially considering the variation in topology with outages and the inherent graph structure of the DN with correlation among different nodal variables.

Persistent Homology Persistent homology [27], [28] is a suite of tools within topological data analysis (TDA) that has shown substantial promise in a broad range of study domains, from bioinformatics to material science to social networks to energy systems [29], [30]. Persistent homology (PH) has also been successfully integrated as a fully trainable topological layer into various deep learning models, addressing tasks such as node and graph classification, link prediction, and anomaly detection (see, e.g., overviews in [30], [31]). More specifically, the goal of PH is to study properties of observed data that are invariant under continuous transformations such as twisting, stretching, and compressing. Such properties are

broadly referred to as *shape*, or *topological* characteristics of the data. Nevertheless, persistent homology and TDA have not yet been employed for resilience related tasks in power distribution systems.

Simplicial Neural Networks Modeling higher-order interactions on graphs is an emerging direction in graph representation learning. While the role of higher-order structures for graph learning has been documented for a number of years [32], [33] and involves diverse applications such as graph signal processing in electricity networks [34], [35], dynamics of disease transmission and biological networks, the integration of higher-order graph substructures into deep learning on graphs has emerged only in 2020. As shown in [36], [37], higher-order network structures can be leveraged to improve the performance of link and trajectory prediction tasks. Indeed, several recent studies [38]–[41] propose to leverage simplicial information to train neural networks on graphs. However, none of these so-called Simplicial Neural Networks (SNNs) is integrated with a fully trainable topological layer allowing them to learn both persistent topological features and simplicial geometry of graphs. In this paper, HOT-Nets is proposed to address this limitation.

Graph Classification with Graph Neural Networks The field of applying graph neural networks for graph classification has received considerable attention in recent years. Graph Convolutional Networks (GCNs) [42] employs a simplified graph convolution by aggregating the target node feature information from its neighbors. GraphSage [43] performs sum, average, or max-pooling neighborhood aggregation and updates the node representation by applying a linear transformation on top of graph convolution. Graph Attention Networks (GAT) [44] combines graph neural networks with attention mechanism by computing attention scores using each node's features and those of its one-hop neighbors. Adaptive Multi-channel Graph Convolutional Networks (AM-GCN) [45] extract different embeddings from node features, topological structures and use an attention mechanism to learn the importance weights for different embeddings. Moreover, some graph neural networks (GNN)-based models exploit a pooling architecture which addresses the limitations of traditional graph pooling architectures, i.e., being incapable of capturing the graph substructure information. For instance, DiffPool [46] develops a differential pooling operator that learns a soft assignment at each graph convolutional layer. A similar idea is utilized in EigenGCN [47], which introduces a pooling operator based on the graph Fourier transform. Compared with these approaches, our proposed HOT-Nets architecture can capture both the most characteristic topological features of the graph and complex interactions among higher-order graph structures described via simplices. As such, HOT-Nets is the first architecture within geometric deep learning that combines a topological layer with learning on simplices.

III. METHODOLOGY

In this section, we present our proposed HOT-Nets model for the graph classification task. The failure of lines in the power DN affects the network connectivity and may result in the isolation of buses from the main grid. Fault isolation

by the DN protection system may disintegrate it into islands with no power generating resource, resulting in disruption of supply at the load buses, thereby creating a network outage. A line failure in the equivalent graph representation of the DN corresponds to the removal of the particular edge. Despite the removal of some edges (lines), the network may still be functioning. This can be attributed to the meshed structure of some networks and the presence of zero injection nodes in the DN which carry no load. However, the power flow variables are also required to provide an accurate description of the system state. Further, any instance with unserved energy is considered as an outage and those with no loss of energy indicate the converse. This detection task is formulated as a graph classification problem. Fig. 1 shows the overall framework of our proposed model. The key idea is that HOT-Nets incorporates both higher-order structures and local topological information about the distribution network into the neural network architecture. The attention mechanism then captures dependencies among different topological representations explicitly. We start from outlining the HOT-Nets approach and show how the proposed framework is used to learn higher-order representations and topological signatures of distribution networks.

Problem Definition We consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ as a model of a distribution system with nodes \mathcal{V} and edges \mathcal{E} . Here $|\mathcal{V}| = N$ is the number of nodes and $|\mathcal{E}| = M$ is the number of edges. Let $A \in \mathbb{R}^{N \times N}$ denote a symmetric adjacency matrix with N nodes, $X_v \in \mathbb{R}^{N \times d_v}$ be a node feature matrix, and $X_e \in \mathbb{R}^{M \times d_e}$ be an edge feature matrix, where d_v and d_e are the dimensions of node and edge features, respectively. Specifically, $A_{uv} = 1$ implies that there exists an edge between nodes (buses) u and v , and $A_{uv} = 0$, otherwise. Let $D \in \mathbb{R}^{N \times N}$ be a diagonal matrix with $D_{uu} = \sum_v A_{uv}$.

The function of the proposed model is to detect the presence of outages in the distribution network. Thus, an outage detector is a binary classifier with '0' representing a normal operation, and '1' representing power disruption. Although the outages result in a change of network connectivity, the base network topology is only known prior to the detection and location of outages. However, the sensor signals superimposed on the graph, i.e., the node variables (bus voltage and active power supplied) and the edge variable (branch flow) vary with each scenario and can be measured. Additionally, some of the edge properties (such as phases, resistance, reactance, etc.) can also be considered as edge features. Outage detection is a graph classification task which consists of viewing the graph (including its node and edge features) as input and predicting its corresponding label. We assume that each graph belongs to one of \mathcal{C} classes (here 2).

A. Preliminaries on Persistent Homology

The key approach of PH is to first associate a graph \mathcal{G} with some filtration: $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \dots \subseteq \mathcal{G}_k = \mathcal{G}$, and then to count various shape patterns of \mathcal{G}_i (where $i = \{1, 2, \dots, k\}$) such as the number of connected components, triangles, and voids, throughout this nested sequence. To make the counting process systematic, we equip each \mathcal{G}_i with a combinatorial object, called an *abstract simplicial complex*.

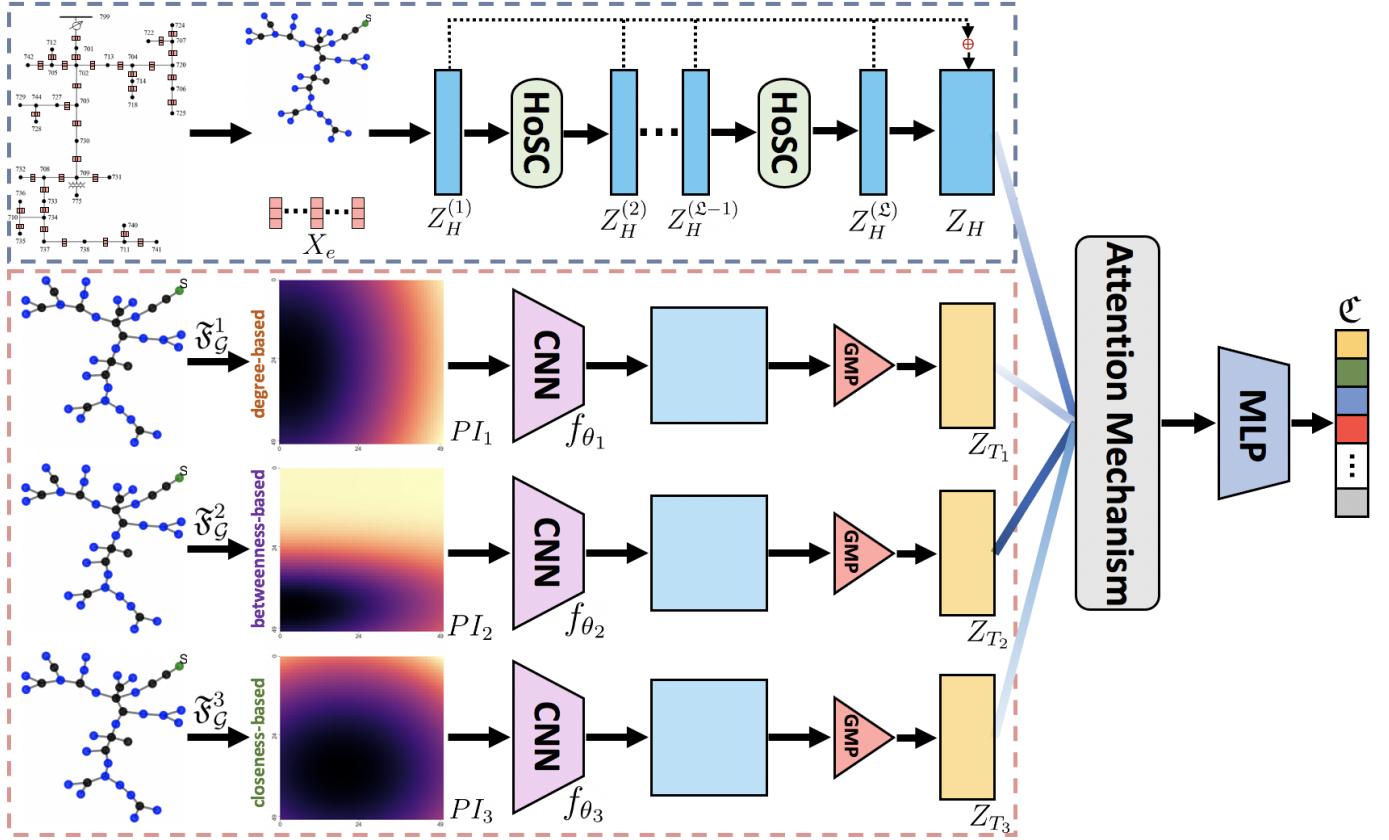


Fig. 1: Framework of our HOT-Nets model for graph classification. *Top row*: The higher-order simplices convolution (HoSC) module is used to extract higher-order simplices embeddings and form a primary higher-order simplex descriptor (i.e., Z_H) via the concatenate operation \oplus . *Bottom row*: First, we generate a persistence image PI_i for the input graph using the filtration \mathfrak{F}_G^i (where $i = \{1, 2, 3\}$, i.e., here we display 3 different filtrations including degree-based, betweenness-based, and closeness-based filtrations); we then feed these PI_i s into a CNN based model to obtain the image-level topological features. An attention mechanism is used to adaptively learn the correlation information among higher-order structures and different topological representations.

Definition 3.1: A family of sets \mathcal{K} with a collection of subsets S is an *abstract simplicial complex* if for every $\sigma \in S$ and every non-empty subset $\tau \subseteq \sigma$, we have $\tau \in \mathcal{K}$, that is, \mathcal{K} is closed under the operation of taking subsets. Elements of \mathcal{K} are called *simplices*. An element $\sigma \in \mathcal{K}$ such that $|\sigma| = k + 1$, is called a *k-simplex*. Every subset $\tau \subset \sigma$ such that $|\tau| = k$ is called a *face* of σ . All simplices in \mathcal{K} that have σ as a face are called *co-faces*. Finally, the dimension d of \mathcal{K} is the largest dimension of any of its faces.

As a result, we represent the graph filtration $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \dots \subseteq \mathcal{G}_k = \mathcal{G}$ as a filtration of abstract simplicial complexes $\mathcal{K}(\mathcal{G}_1) \subseteq \mathcal{K}(\mathcal{G}_2) \subseteq \dots \subseteq \mathcal{K}(\mathcal{G}_k)$, which allows us to use tools of simplicial homology, offering us with computational techniques to study shape characteristics of \mathcal{G} in an efficient manner. In particular, we track which shape patterns appear in this filtration of complexes and record indices of the first and last appearance of each topological feature (i.e., its birth i_b and death i_d , respectively). All extracted topological information can be then summarized in a form of *persistent diagram* (PD) which is a multi-set \mathcal{D} of points in \mathbb{R}^2 , such that the x coordinate is the birth (i.e., i_b) of the q -dimensional topological feature ($0 \leq q \leq d$) and the y coordinate is the death (i.e., i_d) of this topological feature. Since $i_d \geq i_b$, all points in \mathcal{D} are in the half-space on or above $y = x$. *Lifespan*, or *persistence* of a q -dimensional topological feature

is defined as $i_d - i_b$. The longer the lifespan, the likelier the topological feature contains some important information about the structural organization of \mathcal{G} [48], [49]. Features with shorter lifespans are often referred to as *topological noise*.

In general, there are multiple ways that a graph filtration can be constructed (see, e.g., [50]). Here we consider a sublevel filtration induced by a continuous function \mathfrak{F} defined on nodes of \mathcal{G} . That is, let $\mathfrak{F} : \mathcal{V} \rightarrow \mathbb{R}$ and $\nu_1 < \nu_2 < \dots < \nu_n$ be a sequence of sorted filtered values, then $\mathcal{K}_i = \{\sigma \in \mathcal{K} : \max_{v \in \sigma} \mathfrak{F}(v) \leq \nu_i\}$. (A filtration on edges of \mathcal{G} can be defined in a similar manner.) As \mathcal{K} , we use a Vietoris-Rips abstract simplicial complex [51] due to its computational benefits. In our study, we consider \mathfrak{F} to be a function of node degree, betweenness, and closeness. Such an architecture with multiple types of filtration functions allows us to better learn multi-scale network properties along different geometric dimensions. In particular, since in practice the power distribution network is designed to deliver power from the substation to the consumers connected at the different load buses, metrics such as degree and betweenness can identify buses that carry maximal power in the network. In turn, node closeness can identify buses that connect the loads to the substation and hence result in maximum disruption in case of outages.

To encode topological information presented in PD \mathcal{D} into GNNs, we use its vectorized representation, i.e., *persistence*

image (PI) [52]. The PI is defined as a finite-dimensional vector representation derived by a weighted kernel density function and can be formulated in two steps. First, we map \mathcal{D} to an integrable function $\rho_{\mathcal{D}} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, which is called a *persistence surface*. The persistence surface $\rho_{\mathcal{D}}$ is given by sums of weighted Gaussian kernels that are centered at each point in \mathcal{D} . Second, we integrate the persistence surface $\rho_{\mathcal{D}}$ over each grid box to obtain PI.

More specifically, the value of each pixel z within the PI is formed as

$$\text{PI}(z) = \iint \sum_{\mu \in T(\mathcal{D})} \frac{g(\mu)}{2\pi\delta_x\delta_y} e^{-\left(\frac{(x-\mu_x)^2}{2\delta_x^2} + \frac{(y-\mu_y)^2}{2\delta_y^2}\right)} dy dx, \quad (1)$$

where $T(\mathcal{D})$ is the transformation of \mathcal{D} (i.e., $T(x, y) = (x, y - x)$), $g(\mu)$ is a weighting function (where the mean $\mu = (\mu_x, \mu_y) \in \mathbb{R}^2$), and δ_x and δ_y are the standard deviations of the Gaussian kernels in the x and y direction. To gain a better understanding of the complex representations of graph data, we consider multiple types of filtration functions and each filtration function corresponds to a persistence image.

B. Higher-order Simplices Convolution Module

A closely related concept to simplicial homology is the Hodge theory, allowing us to extend the notion of a standard combinatorial graph Laplacian, which addresses diffusion from node to node of \mathcal{G} through edges, to diffusion over higher-order substructures of \mathcal{G} , described by k -simplices of \mathcal{G} . The generalization of graph Laplacians enables us to account for complex multi-node interactions in \mathcal{G} , beyond the node level [36], [53]. Such higher-order interactions are particularly important in the analysis of DN, where the buses (corresponding to nodes in the graph) interact with one another and the power flow through the branches or edges is a result of this interaction. Any change introduced at the buses, such as variation in load demand or power generation, affects the state of all the system variables such as voltage, current, etc.

Definition 3.2: Let C^k be a real-valued vector space endowed with a basis from the oriented k -simplices. (By orientation of simplices, we mean selecting some (arbitrary) order for its nodes, where two orderings are said to be equivalent if they differ by an even permutation.) A linear map $\partial_k : C^k \rightarrow C^{k-1}$ is called a *boundary operator*. The adjoint of the boundary map induces the *co-boundary operator* $\partial_k^T : C^k \rightarrow C^{k+1}$. Matrix representations of ∂_k and ∂_k^T are B_k and B_k^T , respectively. In Fig. 2, we illustrate an example of generating simplicial complexes from a synthetic power network.

An operator over oriented k -simplices $L_k : C^k \rightarrow C^k$ is called the k -*Hodge Laplacian*, and its matrix representation is given by

$$L_k = B_k^T B_k + B_{k+1} B_{k+1}^T, \quad (2)$$

where $B_k^T B_k$ and $B_{k+1} B_{k+1}^T$ are often referred to L_k^{down} and L_k^{up} , respectively.

For example, the standard graph Laplacian $L_0 = B_1 B_1^T \in \mathbb{R}^{N \times N}$ is a special case of the above k -th combinatorial Hodge Laplacian and the matrix $L_1 \in \mathbb{R}^{M \times M}$ is the Hodge 1-Laplacian. In our experiments, following the spectral prop-

erties of the normalized Laplacian operator (see Schaub et al. [37]), we consider the normalized Hodge 1-Laplacian

$$\tilde{L}_1 = D_2 B_1^T D_1^{-1} B_1 + B_2 D_3 B_2^T D_2^{-1}, \quad (3)$$

where $D_2 = \max(\text{diag}(|B_2| \mathbb{1}), I)$ represents a diagonal degree matrix of the edges, $D_1 = 2\text{diag}(|B_1| D_2 \mathbb{1})$ represents a diagonal degree matrix of the nodes, and $D_3 = \frac{1}{3} \mathbb{1}$. For the sake of simplicity and without loss of generality, we set $k = 1$. Furthermore, we define the propagation rule for the normalized Hodge 1-Laplacian, i.e., higher-order simplices convolution (HoSC) module as follows:

$$Z_H^{(\ell+1)} = \max(\Psi(\tilde{L}_1 Z_H^{(\ell)} \Theta_H^{(\ell)})), \quad (4)$$

where $Z_H^{(\ell)} \in \mathbb{R}^{M \times d_{\ell}}$ is the input activation matrix to the ℓ -th hidden layer (where $Z_H^{(0)} = X_e$), $\Theta_H^{(\ell)} \in \mathbb{R}^{d_{\ell} \times d_{\ell+1}}$ is the ℓ -th layer trainable weight matrix, $\max(\cdot)$ denotes the element-wise max operator, and $\Psi(\cdot)$ is a non-linear transform consisting of a Batch Normalization followed by a ReLU activation. In this case, the message passing scheme on simplicial complexes are (i) from edges to edges through nodes and (ii) from edges to edges through triangles – i.e., capturing how information from the edge propagates through the surrounding nodes and faces.

The power distribution network is a real-world structure with complex interdependencies among its key constituents such as power flow through branches and the load (or generation) at the buses. Hence, the state of the power grid network can be described by properties of buses (node features), flow through branches (edge features) and network topology. Encompassing all these within the same framework is pertinent for realistic decision making as opposed to conceptual graphs where one category of features is more relevant than others. Hence, we use a Hodge 1-Laplacian considering the electrical and topological properties of the distribution grid. From a deep learning perspective, all learned higher-order simplices embeddings in different levels (layers) are combined to be a primary higher-order simplices descriptor. That is, we concatenate the

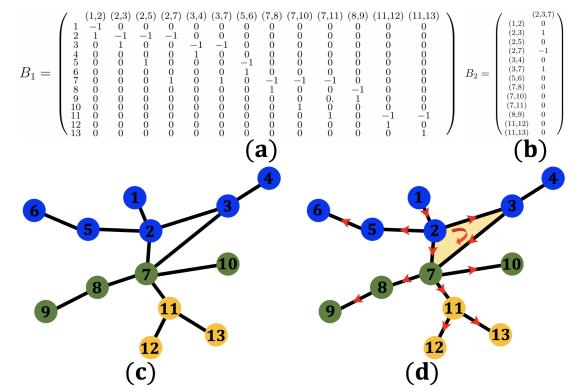


Fig. 2: An example of generating simplicial complexes from a synthetic power network, where each node represents a bus. (a) B_1 is the node-to-edge incidence matrix; (b) B_2 is the edge-to-face incidence matrix; (c) the graph structure of the synthetic power network; (d) the simplicial complexes of the synthetic power network.

output of the \mathcal{L} higher-order simplices convolution modules along the column dimension as

$$Z_H = \oplus(Z_H^{(1)}, \dots, Z_H^{(\mathcal{L})}), \quad (5)$$

where \oplus denotes concatenate operation.

C. Topological Signatures Representation Learning

In our experiments, we use a convolutional neural network (CNN) based model to learn the topological features of a persistence image. Given the persistence image of resolution p , i.e., $PI_i \in \mathbb{R}^{p \times p}$ (where $i = \{1, \dots, \xi\}$ and ξ represents the number of different types of filtrations), we employ a CNN based model and global max pooling to obtain the image-level local topological feature Z_{T_i} as

$$Z_{T_i} = f_{\text{GMP}}(f_{\theta_i}(PI_i)), \quad (6)$$

where f_{GMP} is the global max pooling, f_{θ_i} is a CNN based neural network with parameter set θ_i , and $Z_{T_i} \in \mathbb{R}^{d_c}$ is the output for the i -th persistence image PI_i . Hence, we obtain ξ different learned topological representations $\llbracket Z_{T_1}, \dots, Z_{T_\xi} \rrbracket$ of a graph, which enables us to capture graph properties along various geometric dimensions.

D. Attention Mechanism

To adaptively learn the intrinsic dependencies among different higher-order simplices and topological representations, we utilize the attention mechanism to focus on the importance of task relevant parts of the learned representations for decision making, i.e., $(\alpha_H, \alpha_{T_1}, \dots, \alpha_{T_\xi}) = \text{Att}(Z_H, Z_{T_1}, \dots, Z_{T_\xi})$. In practice, we compute the attention coefficient as follows:

$$\begin{aligned} \alpha_i &= \text{softmax}_i(\Upsilon_{\text{Att}} \tanh(\Xi Z_i)) \\ &= \frac{\exp(\Upsilon_{\text{Att}} \tanh(\Xi Z_i))}{\sum_{j \in \{H, T_1, \dots, T_\xi\}} \exp(\Upsilon_{\text{Att}} \tanh(\Xi Z_j))}, \end{aligned} \quad (7)$$

where $\Upsilon_{\text{Att}} \in \mathbb{R}^{1 \times d_{\text{out}}}$ is a linear transformation, Ξ is the trainable weight matrix, and the softmax function is used to normalize the attention vector. Then, we obtain the final embedding Z by combining all embeddings

$$Z = \alpha_H \times Z_H + \alpha_{T_1} \times Z_{T_1} + \dots + \alpha_{T_\xi} \times Z_{T_\xi}. \quad (8)$$

Lastly, we feed the final embedding Z into a multilayer perceptron (MLP) layer and use a differentiable classifier (here we use a softmax layer) to make graph classification.

IV. EXPERIMENTS

A. Test Networks

The HOT-Nets model is validated on three different distribution test networks, namely the IEEE 37-bus, IEEE 123-bus, and 342-bus low voltage networks [54]. The IEEE 37-bus distribution network is a small size, three-phase, unbalanced, and delta configured medium voltage network rated at 4.8 kV. The 123-bus network is also a three-phase, unbalanced system that operates at 4.16 kV, and has multiple shunt capacitors and voltage regulators. The 342-bus low voltage network (LVN) on the other hand, is a moderate size urban network that has a meshed structure found in North America and is unlike the traditional radial feeders. The total load on the 342-bus LVN is approximately 50 MVA and is supplied by

a 230 kV substation. The power flow and the state for the test networks with varying scenarios are evaluated using the OpenDSS simulation software [55].

B. Dataset Generation

Due to the lack of available data comprising of signal measurements for different outage scenarios, a synthetic approach is used to train the HOT-Nets model. This includes a scenario generation method and evaluation of network behavior by implementing the scenarios on the equivalent DN model in the OpenDSS simulation tool. The network state in each scenario includes measurements such as bus voltage, power supplied at the buses and branch flows through the lines or transformers. These measurements are extracted only for those buses or lines which have a sensor installed so as to emulate a practical distribution network. The buses and lines with existing sensors are assumed in our study whereas this is available information in the real-world DNs.

The scenario generation method accounts for variations in line failures and load profiles. This also encompasses scenarios with normal operating conditions but with load variations to be able to differentiate the outage from normal operation. The degradation of the distribution system due to component failures can be approximated by randomly disconnecting components from the network [56]. A random edge or component removal approach can be used to obtain a generalized model which is independent of the type of the disturbance or the specificity of the weather-related event causing the outage in the feeder. The nature of outages in the distribution network, however, is often localized with a cascading effect. To emulate the behavior of degradation of a real-world distribution feeder, we adopt a subgraph-based random edge removal as in [57]. For the graph model representing the DN, considering the localized effect of contingency events, a subgraph of radius r^s can be initiated around a node u_i selected at random. Following this, a fraction of edges f^s are randomly removed within this subgraph. Since the distribution network tends to be geographically constrained, the probability of failure of the edges due to extreme events can be drawn from a uniform random distribution. The radius of the subgraph r^s , a fraction of edge failure f^s , and the central node u_i are varied to obtain multiple scenarios. Additionally, annual historical load data with an hourly resolution is used to account for the variation in load in the DN. The time of outage for scenarios with line failures is drawn randomly from a uniform distribution, and the corresponding load profile is used in each scenario. In situations where outage data for weather-related events are available, contingency scenarios could also be generated by simulating edge failures drawn from probability distributions derived from the historical data. However, such a Bayesian inference approach is beyond the scope of this work. Equivalent graphs of the distribution feeders are constructed by representing the substation, distribution transformers, and other interconnecting buses as the nodes. The distribution lines and primary transformers in the network are considered as the edges.

Although the line failures are known while generating scenarios for training, in the real-world network operation, this

information is not known a-priori. Rather, the purpose of our model is to detect the presence of one such failure. Relying on the knowledge of the fully functional network topology, and the varying node and edge attributes, the outage detection model determines if edges (lines) have failed, and thus detects the topology change due to line failures. Therefore, the graph scenarios are dynamic in that the signals associated with nodes and edges are varying. Note that, our topological features (e.g., persistence images) are extracted by using various filtration functions (on line features and bus features) and thus the topological signature is varying with nodes and edges in a dynamic system.

We validate the effectiveness of our model using two separate case studies, resulting in two different datasets. In the first case, we assume that the actual power supplied at all the buses can be measured by the sensors. Hence, the active power supplied and demand estimate are considered as the descriptive node features. The edges, however, do not have sensors, and hence different properties (such as resistance, reactance, phase interconnections, maximum capacity rating, baseload and residual capacity) are used as representative edge features.

The following are the properties used to describe an edge (u, v) :

- The phase interconnection.
- The resistance $\mathcal{R}_{(u,v)}$ is the equivalent line resistance evaluated from the three phase matrix of each DN branch.
- The reactance $\mathcal{X}_{(u,v)}$ is the equivalent line reactance evaluated from the three phase matrix of each DN branch.
- The base load capacity $\mathcal{I}_{(u,v)}^B$ is the flow through the distribution line or transformer for base load condition without contingency.
- The maximum capacity $\mathcal{I}_{(u,v)}^{max}$ is the maximum permissible flow through the line or transformer.
- The residual capacity $\mathcal{I}_{(u,v)}^{res}$ is $(|\mathcal{I}_{(u,v)}^{max} - \mathcal{I}_{(u,v)}^B|)$.

A statistical overview of the dataset for all three networks in the first case study is given in Table I.

For the second case, we assume partial observability in the network with sensors placed at 45-50% of the buses and lines in the network. Here the node features include the three-phase voltage, active power supplied, and the demand estimate at the buses. Of these, the former two measurements are available in real-time only at the nodes with sensors placed. On the other hand, the demand estimate, i.e., the demand forecasted at the load buses is a pseudo measurement. The edge features include resistance, reactance, maximum capacity, residual capacity at base load, and power flow through the branch. Except for the branchflow, the others are properties available to the distribution system operator for all edges. The branchflow, on the other hand, is measured continuously only at those edges with line sensors and varies with outages. In the synthetic approach that we adopt, the sensor measurements are acquired from the powerflow simulator. Table II summarizes the statistics of the dataset for all three networks considered in the study with partial observability.

A graphical representation of the 342-bus LVN used in the study is shown in Fig. 3. A representative failure scenario in the 342-bus LVN is illustrated in Fig. 4. Similarly, the IEEE

37-bus network along with a contingency event induced on the network is represented in Fig. 5.

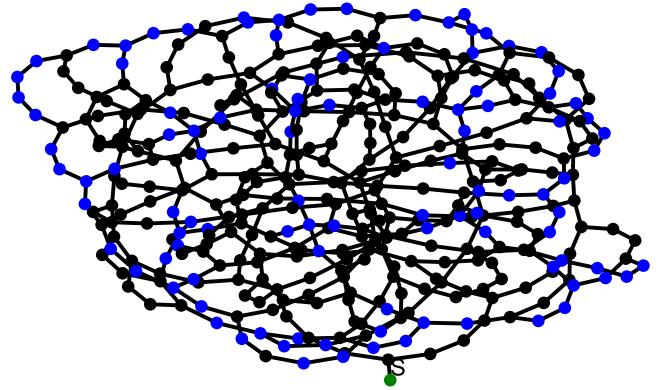


Fig. 3: Graphical representation of the 342-bus low voltage North American distribution network. The source node is marked 'S' and represented in green, blue nodes are buses with loads, and black nodes are the interconnecting buses.

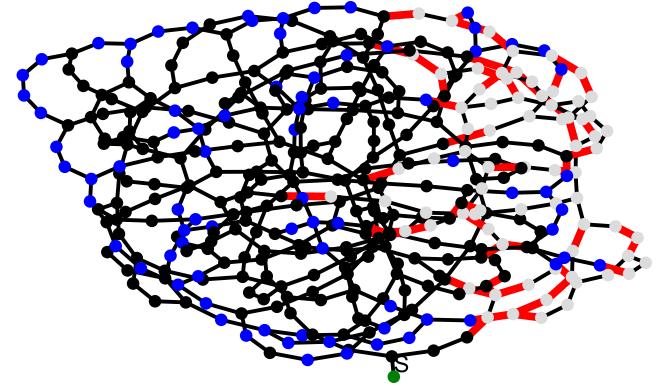


Fig. 4: Illustration of a contingency event on the 342-bus low voltage network. Nodes and edges are color coded. Grey nodes are buses isolated by the network failure, black edges are functioning lines or transformers, and the failed components or edges are represented in red.

TABLE I
SUMMARY OF DATASETS USED IN GRAPH CLASSIFICATION TASK WITH FULL OBSERVABILITY AT BUSES.

Dataset	Graphs	Nodes	Edges [†]	Features _V	Features _E	Classes
IEEE 37 Bus	200	39	35.34	2	8	2
IEEE 123 Bus	300	132	126.56	2	8	2
342 Bus LVN	500	390	432.39	2	8	2

The $[\dagger]$ means the average number of edges in a distribution network under contingency (edge failed).

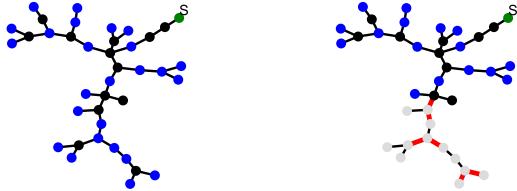
C. Baselines

We compare our proposed HOT-Nets with 11 state-of-the-art (SOA) models, including (i) Random Forest (RF) [58], which is a combination of a series of tree structure classifiers, (ii) Artificial Neural Networks (ANN), which is a feedforward multilayer perceptron (MLP) and can capture both linear and nonlinear features; (iii) GCN, which learns node representations by aggregating representations from neighbors; (iv) GAT [59], which extends the graph convolutional operations

TABLE II
SUMMARY OF DATASETS USED IN GRAPH CLASSIFICATION TASK WITH PARTIALLY OBSERVABLE BUSES AND LINES.

Dataset	Graphs	Nodes	Edges [†]	Features _V	Features _E	Classes
IEEE 37 Bus'	300	39	37.34	5	6	2
IEEE 123 Bus'	300	132	122.87	5	6	2
342 Bus LVN'	300	390	446.74	5	6	2

The $[\dagger]$ means the average number of edges in distribution network under contingency (edge failed).



(a) Base network without contingency. (b) Network with a contingency scenario.

Fig. 5: Graphical representation of the IEEE 37-bus distribution network and an illustration of a contingency event on the network. The source node is represented in green and marked 'S', the blue nodes indicate load buses and the black nodes are interconnecting buses. The failed components or edges are marked in red and grey nodes are buses isolated by the failure.

in GCN with masked self-attentional layers; (v) Graph Iso-morphism Network (GIN) [60], whose representation power is well-matched with the Weisfeiler-Lehman test for graph isomorphism; (vi) GraphSage [43], a graph convolutional network framework that proposes different types of aggregator functions; (vii) Set2Set [61], a graph convolutional network framework, which further replaces the global mean-pooling by global pooling operation through Long Short-Term Memory (LSTM) networks; (viii) DiffPool [46], a graph convolutional network model designed for graph level representation learning with differential pooling layers which coarsens the graph in a hierarchical manner; (ix) EigenGCN [47], a graph convolutional network that deploys pooling operator based on the graph Fourier transform; (x) Adaptive Multi-channel Graph Convolutional Networks (AM-GCN) [45], an adaptive multi-channel graph convolutional network model that learns the node embedding based on both node features and topological structures with attention mechanism; (xi) SNNs [38], a convolutional neural architecture for data supported on simplicial complexes.

D. Experimental Setups

All the SOAs, including our HOT-Nets, are implemented in Python with Pytorch 1.8.0 and executed on a server with one NVIDIA RTX 3090 GPU card. We optimize all the models by the Adam optimizer for a maximum of 50 epochs. We conduct our experiments using 10-fold cross-validation and report the average accuracy. The learning rate is searched in $\{0.1, 0.01, 0.001, 1e-4, 1e-5\}$ and layers of higher-order simplices convolution module $\mathcal{L} \in \{1, 2, \dots, 5\}$ with the hidden layer dimension $nhid^{\text{HoSC}} \in \{8, 16, 32, 64\}$, and the dropout rate is 0. For PI representation learning (i.e., PIs from 3 different filtrations), we train three 2-layer CNN based models with the same hidden layer dimension $nhid_1^{\text{CNN}} \in \{8, 16, 32\}$ and the same output dimension $nhid_2^{\text{CNN}} \in \{16, 32, 64\}$ simultaneously. Energy is said to be unserved at a node if

the estimated load demand at the particular node is not met. The total energy unserved for the network is the sum of the energy unserved at all the nodes and it indicates an outage in the network. The opposite is true if there is no unserved energy. Hence, we label each network scenario in the outage detection task, thereby resulting in a graph classification with 2 classes. In our experiments, we consider $\xi = 3$ different filtrations (i.e., degree-based, betweenness-based, and closeness-based filtrations via setting the grid size of the PIs to 50×50) and note that the proposed topological signatures representation learning module can learn an arbitrary number of filtrations on the graph. The source code is available at <https://github.com/hotnets/HOT-Nets.git>.

E. Graph Classification

The evaluation results are summarized in Table III. In particular, on all three networks, the improvement gain of HOT-Nets over the runner-ups ranges from 2.88% to 7.63%. In terms of baseline methods, GCN only takes information from 1-hop neighbors into consideration and is not capable of learning topological representations of a graph. GraphSage learns a principle of aggregation to extend GCN into the inductive setting and shows stable improvement over GCN. Besides, Set2Set, DiffPool, EigenGCN, and AM-GCN are SOA GCN-based models (including global pooling, hierarchical pooling, attention mechanism implemented over a graph network architecture, or an extension of GCN architecture to process data supported on simplicial complexes), and SNNs are SNN-based model. A common limitation of these SOA methods is that they are incapable of incorporating both higher-order features and multi-scale local topological structures. Therefore, it is natural that HOT-Nets shows much better performance against these SOA baselines.

The results on the case study with partially observable test networks (i.e., IEEE 37 Bus', IEEE 123 Bus', and 342 Bus LVN') are summarized in Table IV. For IEEE 37 Bus', we observe that our HOT-Nets yields significant relative gains of 2.37% on average compared to all 9 baselines. Specifically, HOT-Nets surpasses the performances of GNN-based models by a notable margin to reach 5.36% and outperforms the simplicial complex-based model (SNNs) by 3.09%. For IEEE 123 Bus', we find that our HOT-Nets model significantly outperforms all baselines, i.e., $p\text{-value} \ll 0.01$ by t-test. Specifically, HOT-Nets can improve upon the runner-up of attention-based GNNs (i.e., AM-GCN) and simplicial complex-based model (i.e., SNNs) by a margin of 2.54 and 2.36 on IEEE 123-Bus' dataset (i.e., with unobservable distribution grids scenario), respectively. Furthermore, it is apparent to see that, our proposed HOT-Nets significantly outperforms all baselines. HOT-Nets brings more than 4% relative improvement to the runner-up (i.e., DiffPool) in accuracy on 342 Bus LVN'.

Furthermore, we have also considered that real-time measurements are acquired from the node and line sensors placed in the network as in [5]. Taking into account the suggestion of the reviewer, we have conducted another experiment for 123 bus network (denoted by IEEE 123 Bus'') with 30% sensor density which is also used in [5]. The results of this experiment

TABLE III
OVERALL CLASSIFICATION PERFORMANCE (%) (\pm STANDARD DEVIATION) OF DIFFERENT METHODS ON TEST NETWORKS FOR A CASE WITH ALL BUSES OBSERVABLE. *** DENOTES THE HIGHLY STATISTICALLY SIGNIFICANT RESULT.

Datasets	RF	ANN	GCN	GAT	GIN	GraphSage	Set2Set	DiffPool	EigenGCN	AM-GCN	SNNs	HOT-Nets (ours)
IEEE 37 Bus	77.61 \pm 1.55	78.21 \pm 1.48	84.45 \pm 1.67	85.02 \pm 1.74	87.28 \pm 1.97	86.96 \pm 2.25	88.20 \pm 1.94	90.25 \pm 2.32	87.50 \pm 2.36	83.98 \pm 2.03	86.65 \pm 1.80	***97.70 \pm 1.64
IEEE 123 Bus	68.50 \pm 0.25	73.15 \pm 0.85	84.33 \pm 0.77	83.80 \pm 0.76	81.91 \pm 0.79	86.00 \pm 1.30	76.67 \pm 0.82	87.67 \pm 0.90	87.73 \pm 0.89	87.68 \pm 0.15	86.69 \pm 1.01	***90.33 \pm 0.73
342-bus LVN	63.32 \pm 0.18	68.07 \pm 0.72	75.05 \pm 1.70	73.62 \pm 1.98	74.80 \pm 1.91	76.47 \pm 1.67	78.39 \pm 1.79	77.93 \pm 1.22	80.61 \pm 1.97	76.20 \pm 2.00	79.26 \pm 1.31	***83.68 \pm 2.03

TABLE IV
CLASSIFICATION PERFORMANCE (%) (\pm STANDARD DEVIATION) OF DIFFERENT METHODS ON IEEE 37 Bus', IEEE 123 Bus', AND 342 Bus LVN' (I.E., NEW TEST CASE ON NETWORKS) WITH PARTIALLY OBSERVABLE DISTRIBUTION GRIDS. *** DENOTES THE HIGHLY STATISTICALLY SIGNIFICANT RESULT.

Datasets	RF	ANN	GCN	GAT	GIN	GraphSage	Set2Set	DiffPool	EigenGCN	AM-GCN	SNNs	HOT-Nets (ours)
IEEE 37 Bus'	83.60 \pm 0.16	85.28 \pm 0.49	91.96 \pm 0.37	93.82 \pm 0.25	92.81 \pm 0.63	93.70 \pm 0.50	92.00 \pm 0.26	94.60 \pm 0.40	93.66 \pm 0.23	94.24 \pm 0.25	94.42 \pm 0.19	***95.67 \pm 0.15
IEEE 123 Bus'	84.29 \pm 0.30	87.26 \pm 0.92	93.50 \pm 0.50	94.44 \pm 0.75	94.62 \pm 0.73	95.03 \pm 0.37	91.43 \pm 0.22	95.29 \pm 0.43	94.33 \pm 0.86	95.12 \pm 0.66	95.30 \pm 0.43	***97.66 \pm 0.33
342 Bus LVN'	76.49 \pm 0.20	76.57 \pm 0.30	83.11 \pm 0.34	83.95 \pm 0.46	85.44 \pm 0.52	84.67 \pm 0.59	83.95 \pm 0.93	87.24 \pm 0.78	83.76 \pm 0.50	85.95 \pm 0.61	89.47 \pm 0.17	***91.11 \pm 0.46

is presented in Table V. We observe that our HOT-Nets clearly outperforms all baselines, leading to a relative gain of 3.57% compared to the runner-up (i.e., SNNs). As with the case study for 50% sensor density, here the sensors have been randomly pre-allocated at specific buses and lines. The node features and edge features remain the same as with previous studies. The voltage, power supplied at buses, and the power flow through branches are available in real-time only for those nodes and edges with sensors.

F. Ablation Study

To better evaluate the performance of Hodge Laplacian representation, persistence images from multiple filtrations (i.e., multifiltrations), and attention mechanism, we conduct a comprehensive ablation study and the results are presented in Table VI. The results show that HOT-Nets has better performance over NT-Nets (i.e., instead of using the normalized Hodge 1-Laplacian, NT-Nets uses the normalized node-level Laplacian), ET-Nets (i.e., instead of using the normalized Hodge 1-Laplacian, ET-Nets uses the normalized edge-level Laplacian), HOT-Nets without attention mechanism (HOT-Nets W/o Att.), HOT-Nets with only one PI for topological signatures representation learning (HOT-Nets With One Topo.), or HOT-Nets without any PI for topological signatures representation learning (HOT-Nets W/o Topo.). Specifically, we observe that when replacing the Hodge 1-Laplacian with the node or edge-level Laplacian, the graph classification accuracy is affected significantly, i.e., HOT-Nets outperforms NT-Nets with relative gains of 12.99% and 4.12% for IEEE 37-bus and IEEE 123-bus networks, respectively. Moreover, the results indicate the utility of the proposed attention mechanism and integration of multiple PIs into the representation learning of topological information, in conjunction with the classification of distribution grid status. Furthermore, we find that combining the simplicial convolutional layer and the fully trainable topological layer results in more informative learning of the underlying graph structure, yielding more competitive classification performance (where HOT-Nets achieves relative gains of up to 7.83% and 12.99% over HOT-Nets W/o Topo. and HOT-Nets W/o HoSC (NT-Nets), respectively).

Similarly, to demonstrate the effectiveness of our HOT-Nets model in the IEEE 123 Bus' dataset, we have added

the ablation study. From Table VII, the results show that all the components are indispensable. Specifically, NT-Nets (i.e., without using the normalized Hodge 1-Laplacian) indicates the importance of higher-order structures information for contingency classification of distribution networks. Furthermore, HOT-Nets w/o Topo. (i.e., without any PI for topological signatures representation learning) shows the advantage of leveraging topological summaries to capture hidden structural and local topological information in distribution networks.

Our experiments demonstrate that HOT-Nets, if employed in power distribution networks, can detect outages caused due to single and multi component failures with higher accuracy than SOA graph and simplicial based learning techniques. This provides the much needed situational awareness for the network under contingency, thereby enabling distribution system operators (DSOs) to take suitable actions to mitigate outages. Our model can also be used to assess the risk associated with potential failure scenarios since it predicts the occurrence of an outage. In cases where the component failures do not result in load disruptions, the DSO can simply ensure that the faulty area is isolated and continue to monitor power balance in the functioning network component. However, if the failure event results in power outage, immediate actions following isolation such as dispatch of energy resources, repair crews, and other control & switching actions may be required.

G. Computational Complexity

For higher-order simplices, incidence matrices B_1 and B_2 can be calculated efficiently with computational complexity $\mathcal{O}(N+M)$ and $\mathcal{O}(M+Q)$ respectively, where N is the number of 0-simplices (i.e., nodes), M is the number of 1-simplices (i.e., edges), and Q is the number of 2-simplices (i.e., filled triangles). Persistent homology can be calculated efficiently for dimensions 0 and 1 (i.e., having a worst-case complexity of $\mathcal{O}(M\eta(M))$ with M sorted edges and $\eta(\cdot)$ denotes the extremely slow growing inverse Ackermann function) and the complexity of calculating persistent features from a filtration is dominated by the complexity of sorting all edges, i.e., $\mathcal{O}(M \log M)$.

H. Application to Practical Distribution Networks

Real-world distribution networks are widespread with hundreds of thousands of nodes. Hence, sensor placement at each

TABLE V
OVERALL CLASSIFICATION PERFORMANCE (%) (\pm STANDARD DEVIATION) OF DIFFERENT METHODS ON IEEE 123 BUS'' WITH 30% SENSOR DENSITY. *** DENOTES THE HIGHLY STATISTICALLY SIGNIFICANT RESULT.

Datasets	RF	ANN	GCN	GAT	GIN	GraphSage	Set2Set	DiffPool	EigenGCN	AM-GCN	SNNs	HOT-Nets (ours)
IEEE 123 Bus''	70.43 \pm 0.34	78.23 \pm 0.20	81.53 \pm 0.57	82.37 \pm 0.72	83.92 \pm 0.56	83.75 \pm 0.38	84.53 \pm 0.58	85.20 \pm 0.43	86.36 \pm 0.75	87.69 \pm 0.74	88.11 \pm 0.69	***91.37 \pm 0.67

TABLE VI
ABLATION STUDY OF THE NETWORK ARCHITECTURE.

Method	Architecture			Datasets	
	Hodge	Topo.	Att.	IEEE 37 Bus	IEEE 123 Bus
HOT-Nets	✓	✓	✓	***97.70	***90.33
Node & Topo.-Nets (NT-Nets)	✗	✓	✓	85.00	86.61
Edge & Topo.-Nets (ET-Nets)	✗	✓	✓	86.93	87.71
HOT-Nets W/o Att.	✓	✓	✗	93.49	88.59
HOT-Nets With One Topo.	✓	✗	✓	92.75	89.00
HOT-Nets W/o Topo.	✓	✗	✓	90.05	88.11

TABLE VII
ABLATION STUDY OF THE NETWORK ARCHITECTURE ON IEEE 123-Bus'. * DENOTES THE SIGNIFICANT RESULT.

Method	Architecture			Datasets	
	Hodge	Topo.	Att.	IEEE 123 Bus'	
HOT-Nets	✓	✓	✓	✓	*97.66
Node & Topo.-Nets (NT-Nets)	✗	✓	✓	✓	86.33
Edge & Topo.-Nets (ET-Nets)	✗	✓	✓	✓	96.17
HOT-Nets W/o Att.	✓	✓	✗	✓	94.53
HOT-Nets With One Topo.	✓	✗	✓	✓	95.65
HOT-Nets W/o Topo.	✓	✗	✓	✓	92.35

line or bus is impractical, considering the installation and maintenance cost, and also the communication bottleneck. Therefore, sparse signals collected from a few sensors are to be used to accurately detect the network status in real time. Learning based methods can be used to address these issues. Most traditional learning techniques are however solely data-driven without accounting for the topological interdependence of the complex network. Factoring in topology and also learning from both the global and local structural information is pertinent to derive accurate conclusions about the system state. This is evident from the performance of the HOT-Nets and the ablation study. One such trained model could be deployed in distribution networks where the sparse signal measurements are continuously used to monitor the network status.

V. CONCLUSION

To enable the resilient operation of power distribution systems, we have proposed a new graph learning model that leverages the utility of persistent homology on graphs and extends the convolutional operation to simplicial complexes on the distribution network using Hodge-Laplacian analytics. Integrating persistent homology into learning distribution networks allow us to extract the most characteristic topological descriptors of the distribution grid, while the Hodge-Laplacian analytics account for complex interactions among the higher-

order substructures of the grid. Compared to the state-of-the-art learning methods, the new HOT-Nets model has been shown to deliver highly competitive performance in resilience analysis of power distribution networks by predicting the outage status of a network under varying operating conditions. In the future, we plan to exploit these concepts to also determine the locations of outages and estimate the scale of disruption (energy not served). Additionally, we also plan to explore the advantages and limitations of HOT-Nets and, more generally, simplicial neural networks with a fully trainable topological layer for optimal design problems in general cyber-physical systems.

ACKNOWLEDGMENT

This material is based upon work sponsored in part by the U.S. Department of the Navy, Office of Naval Research under ONR award number N00014-21-1-2530. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research. This work was also sponsored in part by the U.S. National Science Foundation under Grant ECCS-2039716/2039701, and in part by a grant from the C3.ai Digital Transformation Institute. The paper has been awarded the 2022 Best Student Paper Award from the Section for Statistics in Defense and National Security (SDNS) of the American Statistical Association (ASA).

REFERENCES

- [1] A. N. Samudrala, M. H. Amini, S. Kar, and R. S. Blum, "Sensor placement for outage identifiability in power distribution networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 1996–2013, 2019.
- [2] R. A. Jacob, S. Paul, W. Li, S. Chowdhury, Y. R. Gel, and J. Zhang, "Reconfiguring unbalanced distribution networks using reinforcement learning over graphs," in *IEEE Texas Power and Energy Conference*, 2022, pp. 1–6.
- [3] R. A. Jacob and J. Zhang, "Outage management in active distribution networks with distributed energy resources," in *North American Power Symposium*. IEEE, 2021, pp. 1–6.
- [4] C. Ji, Y. Wei, and H. V. Poor, "Resilience of energy infrastructure and services: modeling, data analytics, and metrics," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1354–1366, 2017.
- [5] A. N. Samudrala, M. H. Amini, S. Kar, and R. S. Blum, "Distributed outage detection in power distribution networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5124–5137, 2020.
- [6] R. A. Sevlian, Y. Zhao, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Outage detection using load and line flow measurements in power distribution systems," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2053–2069, 2017.
- [7] Y. Zhao, R. Sevlian, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Outage detection in power distribution networks with optimally-deployed power flow sensors," in *IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.
- [8] S. Soltan, P. Mittal, and H. V. Poor, "Line failure detection after a cyber-physical attack on the grid using bayesian regression," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3758–3768, 2019.

[9] A. Arabali, M. Ghofrani, and M. Farasat, "A new multiple line outage identification formulation using a sparse vector recovery technique," *Electric Power Systems Research*, vol. 142, pp. 237–248, 2017.

[10] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4910–4920, 2019.

[11] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796–4805, 2019.

[12] B. W. Brunton, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Sparse sensor placement optimization for classification," *SIAM Journal on Applied Mathematics*, vol. 76, no. 5, pp. 2099–2122, 2016.

[13] B. Alnajjar, A. N. Samudrala, C. Chen, R. S. Blum, S. Kar, and E. M. Stewart, "Outage detection for distribution networks using limited number of power flow measurements," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 2, pp. 315–324, 2020.

[14] F. C. Trindade, L. F. Ochoa, and W. Freitas, "Data analytics in smart distribution networks: Applications and challenges," *IEEE Innovative Smart Grid Technologies-Asia*, pp. 574–579, 2016.

[15] S. T. Mak and N. Farah, "Synchronizing scada and smart meters operation for advanced smart distribution grid applications," in *IEEE PES Innovative Smart Grid Technologies*. IEEE, 2012, pp. 1–7.

[16] R. A. Jacob and J. Zhang, "Outage management in active distribution network with distributed energy resources," in *North American Power Symposium*. IEEE, 2021, pp. 1–6.

[17] T. R. Fernandes, B. Venkatesh, and M. C. de Almeida, "Distribution system topology identification via efficient milp-based wlav state estimation," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 75–84, 2022.

[18] H. S. Karimi and B. Natarajan, "Joint topology identification and state estimation in unobservable distribution grids," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5299–5309, 2021.

[19] A. Jaech, B. Zhang, M. Ostendorf, and D. S. Kirschen, "Real-time prediction of the duration of distribution system outages," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 773–781, 2018.

[20] A. Arif and Z. Wang, "Distribution network outage data analysis and repair time prediction using deep learning," in *Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems*. IEEE, 2018, pp. 1–6.

[21] D. Cerrai, D. W. Wanik, M. A. E. Bhuiyan, X. Zhang, J. Yang, M. E. Frediani, and E. N. Anagnostou, "Predicting storm outages through new representations of weather and vegetation," *IEEE Access*, vol. 7, pp. 29 639–29 654, 2019.

[22] R. Baembitov, M. Kezunovic, and Z. Obradovic, "Graph embeddings for outage prediction," in *North American Power Symposium*. IEEE, 2021, pp. 1–6.

[23] X. Chen, J. Qiu, L. Reedman, and Z. Y. Dong, "A statistical risk assessment framework for distribution network resilience," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4773–4783, 2019.

[24] H. Sun, Z. Wang, J. Wang, Z. Huang, N. Carrington, and J. Liao, "Data-driven power outage detection by social sensors," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2516–2524, 2016.

[25] Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "Outage detection in partially observable distribution systems using smart meters and generative adversarial networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5418–5430, 2020.

[26] B. Azimian, R. S. Biswas, S. Moshtaghi, A. Pal, L. Tong, and G. Dasarathy, "State and topology estimation for unobservable distribution systems using deep neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.

[27] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," in *Proceedings of the Annual Symposium on Foundations of Computer Science*, 2000, pp. 454–463.

[28] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.

[29] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, "A roadmap for the computation of persistent homology," *EPJ Data Science*, vol. 6, pp. 1–38, 2017.

[30] G. Carlsson, "Topological methods for data modelling," *Nature Reviews Physics*, vol. 2, no. 12, pp. 697–708, 2020.

[31] G. Tauzin, U. Lupo, L. Tunstall, J. Pérez, M. Caorsi, A. Medina-Mardones, A. Dassatti, and K. Hess, "giotto-tda: A topological data analysis toolkit for machine learning and data exploration," *Journal of Machine Learning Research*, vol. 22, pp. 39–1, 2021.

[32] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in *Proceedings of the International Conference on Machine Learning*, 2006, pp. 17–24.

[33] J. L. Johnson and T. Goldring, "Discrete hodge theory on graphs: a tutorial," *Computing in Science & Engineering*, vol. 15, no. 5, pp. 42–55, 2012.

[34] J. P. Roth, "An application of algebraic topology to numerical analysis: On the existence of a solution to the network problem," *Proceedings of the National Academy of Sciences*, p. 518, 1955.

[35] ———, "An application of algebraic topology: Kron's method of tearing," *Quarterly Of applied mathematics*, pp. 1–24, 1959.

[36] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, "Simplicial closure and higher-order link prediction," *Proceedings of the National Academy of Sciences*, vol. 115, pp. E11 221–30, 2018.

[37] M. T. Schaub, A. R. Benson, P. Horn, G. Lippner, and A. Jadbabaie, "Random walks on simplicial complexes and the normalized hodge 1-laplacian," *SIAM Review*, vol. 62, no. 2, pp. 353–391, 2020.

[38] S. Ebli, M. Deferrard, and G. Spreemann, "Simplicial neural networks," in *NeurIPS 2020 Workshop on TDA and Beyond*, 2020.

[39] T. M. Roddenberry, N. Glaze, and S. Segarra, "Principled simplicial neural networks for trajectory prediction," in *Proceedings of the International Conference on Machine Learning*, 2021, pp. 9020–9029.

[40] C. Bodnar, F. Frasca, Y. G. Wang, N. Otter, G. Montúfar, P. Lio, and M. Bronstein, "Weisfeiler and lehman go topological: Message passing simplicial networks," in *Proceedings of the International Conference on Machine Learning*, 2021, pp. 1026–1037.

[41] Y. Chen, Y. R. Gel, and H. V. Poor, "BScNets: Block simplicial complex neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 6, 2022, pp. 6333–6341.

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

[43] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, 2017, pp. 1025–1035.

[44] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *Proceedings of the International Conference on Learning Representations*, 2018.

[45] X. Wang, M. Zhu, D. Bo, P. Cui, C. Shi, and J. Pei, "AM-GCN: Adaptive multi-channel graph convolutional networks," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 1243–1253.

[46] R. Ying, J. You, C. Morris, X. Ren, W. L. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," in *Advances in Neural Information Processing Systems*, 2018, pp. 4805–4815.

[47] Y. Ma, S. Wang, C. C. Aggarwal, and J. Tang, "Graph convolutional networks with eigenpooling," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 723–731.

[48] G. Carlsson, "Topological pattern recognition for point cloud data," *Acta Numerica*, pp. 289–368, 2014.

[49] F. Chazal and B. Michel, "An introduction to topological data analysis: fundamental and practical aspects for data scientists," *Frontiers in Artificial Intelligence*, vol. 4, 2021.

[50] C. Hofer, F. Graf, B. Rieck, M. Niethammer, and R. Kwitt, "Graph filtration learning," in *Proceedings of the International Conference on Machine Learning*, 2020, pp. 4314–4323.

[51] J.-C. Hausmann *et al.*, "On the vietoris-rips complexes and a cohomology theory for metric spaces," *Annals of Mathematics Studies*, vol. 138, pp. 175–188, 1995.

[52] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, "Persistence images: A stable vector representation of persistent homology," *Journal of Machine Learning Research*, vol. 18, 2017.

[53] L. H. Lim, "Hodge laplacians on graphs," *SIAM Review*, vol. 62, pp. 685–715, 2020.

[54] K. Schneider, B. Mather, B. Pal, C. Ten, G. Shirek, H. Zhu, J. Fuller, J. Pereira, L. Ochoa, L. de Araujo, and R. Dugan, "Analytic considerations and design basis for the IEEE distribution test feeders," *IEEE Transactions on Power Systems*, vol. 33, pp. 3181–3188, 2017.

[55] R. C. Dugan, "Reference guide: The open distribution system simulator (OpenDSS)," *Electric Power Research Institute*, 2016.

[56] A. M. Danielsson, "Deep learning for power system restoration," Ph.D. dissertation, 2018.

[57] B. Bush, Y. Chen, D. Ofori-Boateng, and Y. R. Gel, "Topological machine learning methods for power system responses to contingencies," in *Proceedings of the Innovative Applications of Artificial Intelligence Conference*, vol. 35, no. 17, 2021, pp. 15 278–15 285.

- [58] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.
- [59] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proceedings of the International Conference on Learning Representations*, 2018.
- [60] 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.
- [61] O. Vinyals, S. Bengio, and M. Kudlur, "Order matters: Sequence to sequence for sets," *Proceedings of the International Conference on Learning Representations*, 2016.



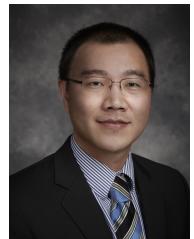
Yuzhou Chen received the M.S. degree in Statistics from the University of Texas at Dallas and the Ph.D. degree in Statistics from Southern Methodist University, in 2017 and 2021. He was a postdoctoral scholar in the Department of Electrical and Computer Engineering at Princeton University in 2021–2022. He is currently an Assistant Professor in the Department of Computer and Information Sciences at Temple University. He is also a Visiting Research Collaborator in the Department of Electrical and Computer Engineering at Princeton University. His research interests include machine learning, deep learning, and topological data analysis with applications to energy systems, health, and social networks.



works and microgrids, resilience enhancement, smart grids, machine/deep learning applications in power grids, and energy management in integrated energy systems.



Yulia R. Gel received the Ph.D. degree in Applied Mathematics from Saint Petersburg State University, Russia in 2000. Following a postdoctoral fellowship in statistics at the University of Washington in 2001–2003, she joined the University of Waterloo, Canada in 2004 where she served as Assistant and then Associate Professor till 2013. Since 2014 she is on the faculty of the University of Texas at Dallas. Currently she is the Program Director at the Division of Mathematical Sciences (DMS) and the Directorate for Technology, Innovation and Partnerships (TIP) of the National Science Foundation of USA. She held visiting positions at Johns Hopkins University, University of California, Berkeley, and the Isaac Newton Institute for Mathematical Sciences, Cambridge University, UK. She is Fellow of the American Statistical Association (ASA). Her research interests include statistical foundation of data science, topological and geometric methods in statistics and machine learning, in application to risk analytics in power systems, environmental studies, biosurveillance, and blockchains.



Jie Zhang (Senior Member, IEEE) received the B.S. and M.S. degrees in mechanical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2006 and 2008, respectively, and the Ph.D. degree in mechanical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2012. He is currently an Associate Professor with the Department of Mechanical Engineering and (Affiliated) Department of Electrical and Computer Engineering, The University of Texas at Dallas. His research interests include integrated energy systems, renewable energy grid integration, complex networks, big data analytics, and multidisciplinary design optimization.



H. Vincent Poor (S'72, M'77, SM'82, F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is currently the Michael Henry Strater University Professor. During 2006 to 2016, he served as the dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other universities, including most recently at Berkeley and Cambridge.

His research interests are in the areas of information theory, machine learning and network science, and their applications in wireless networks, energy systems and related fields. Among his publications in these areas is the book *Advanced Data Analytics for Power Systems*. (Cambridge University Press, 2021). Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences and is a foreign member of the Royal Society and other national and international academies. He received the IEEE Alexander Graham Bell Medal in 2017.