An Overview on the Application of Graph Neural Networks in Wireless Networks

Shiwen He, Member, IEEE, Shaowen Xiong, Yeyu Ou, Jian Zhang Member, IEEE, Jiaheng Wang, Senior Member, IEEE, Yongming Huang, Senior Member, IEEE, and Yaoxue Zhang, Senior Member, IEEE

In recent years, with the rapid enhancement of computing power, deep learning methods have been widely applied in wireless networks and achieved impressive performance. To effectively exploit the information of graph-structured data as well as contextual information, graph neural networks (GNNs) have been introduced to address a series of optimization problems of wireless networks. In this overview, we first illustrate the construction method of wireless communication graph for various wireless networks and simply introduce the progress of several classical paradigms of GNNs. Then, several applications of GNNs in wireless networks such as resource allocation and several emerging fields, are discussed in detail. Finally, some research trends about the applications of GNNs in wireless communication systems are discussed.

Index Terms—Wireless networks, graph neural networks, resource management

I. INTRODUCTION

The advent of fifth-generation (5G) wireless communication systems has driven the revolutionary applications extending far beyond smartphones and other mobile devices [1]. Meanwhile, intelligent communication becomes a novel developing trend of future communication systems [2]. Recently, more and more researchers adopt the deep learning (DL) method to study the problems in wireless networks motivating by the successful application of DL in the related fields of computer [3].

A. From Traditional Deep Learning to Graph Neural Network

According to the usage of domain knowledge, DL methods can be divided into the data-driven DL method and the data- and model-driven DL method. The data-driven DL methods without using the domain knowledge generally have poor interpretability and robustness, while having fast inferencing speed compared with the model-based method. A classic work using multi-layer perceptrons (MLPs) is to solve the power control problem of wireless networks by using the MLPs to approximate the weighted minimum mean square error (WMMSE) algorithm [4]. The data- and model-driven DL methods are rising in recent years and its core idea is to retain the main theoretical characteristics of the classical model algorithms, while using the DL methods to partially enhance or replace its related difficult or time-consuming process. Compared with the data-driven DL methods, the data- and model-driven DL methods have better interpretability and robustness with slower inferencing speed. For example, A. Bora et al. used the generated model from neural networks instead of the standard sparsity model to represent data distributions [5]. N. Shlezinger et al. designed a deep neural network (DNN) to implement the channel-dependent part of Viterbi algorithm, while keeping the unchange of the rest remains [6]. H. Kim et al. studied a family of sequential codes parameterized by recurrent neural network (RNN) architectures [7]. K. Gregor et al. designed a non-linear and parameterized feed-forward architecture with a fixed depth to approximate the optimal sparse code [8]. Some overviews are presented to summary the application of machine learning (ML) or DL with aiming to improve the quality-of-experience (QoE) of wireless networks [9], [10]. The data collected in these tasks is typically represented in the Euclidean domains. Although the existing works using the DL models defined in the Euclidean domains achieve a better performance in small-scale networks, they fail to exploit the underlying topology of wireless networks. Consequently, the performance decreases sharply when the network scale becomes large [11], [12].

In wireless networks, an obvious feature is that the high dynamics of network topology caused by some uncertain factors, e.g., the user mobility, changes in traffic pattern or adjustment of the network resource, etc. In addition, the wireless data may be collected from non-Euclidean domains and represented as graph-structured data with high dimensional features and interdependency among communication devices. These issues bring difficulties to apply directly the learning model defined in Euclidean domains in wireless networks. A straightforward way solving these difficulties is to incorporate the network topology information, which is described as an adjacency matrix depending on the specific node index, into the architecture of neural networks. However, the indices
of communication devices in wireless networks may change due to the reallocation of resources and the movement of communication devices, i.e., dynamic graph-structured data. This motivates us to design a novel learning model with taking into account the interdependencies between communication devices and the dynamics of wireless networks [13]. The emerging GNNs enable the graph-structured data to be processed effectively and to use the global parameterization, common system of coordinates, vector space structure, or shift-invariance [14]. In the last few years, many researchers have begun to use GNNs to mine the deep information hidden in the graph-structured data to further improve the abilities of learning and simulating the interaction between nodes.

B. Related Overviews and Contributions

Some overviews about the paradigms and applications of GNNs are represented in the past few years. The authors of [15] introduced comprehensively four basic paradigms of GNNs and described the representative models in detail. Differentiating from [15], the authors of [16] further reviewed graph reinforcement learning and graph adversarial methods. The authors of [17] summarized the state-of-the-art of the main models and algorithms of graph learning from four aspects, i.e., graph signal processing, matrix decomposition, random walk, and deep learning. The aforementioned surveys introduced in detail the characteristics of different GNN paradigms but briefly introduced the application of GNNs in some fields except for wireless networks. The authors of [18] summarized the GNN approaches in terms of the spatial domain and spectral domain, respectively. Furthermore, this work integrates the spatial and spectral domain models into a unified framework. The authors of [19] further discussed the expressive power of GNNs and further summarized the relationships among GNNs, the Weisfeiler-Lehman algorithm, and distributed local algorithms. The authors of [20] illustrated the excellent performance of GNNs depending on the three characteristics of equivariance, stability and transferability, which are further determined by the selection of the optimization objective and technologies, as well as the design of graph filters. The methods using software or hardware to accelerate GNNs were reviewed from the perspective of computation speed in [21]. In order to fill the research on the interpretability of GNNs, the unified methodology and standard testbed for evaluating the interpretability of GNNs were summarized comprehensively in [22]. There are also several overviews on applying the GNNs to solve the problems in the traffic domain [23], power systems [24], and recommender system [25]. In addition, the relationship between GNNs and the latest neural-symbolic computing that aims at integrating the abilities of learning from the environment and of reasoning from what has been learned was introduced in [26].

In this paper, we aim to present a comprehensive overview of the application of GNNs in wireless networks. Meanwhile, we also provide some potential research directions for researchers who are interested in this topic. A detailed organization of this overview is illustrated in Fig. 1. In particular, the main contributions of this overview are summarized as follow

- The methods of constructing wireless communication graph (WCG) for Mesh/Ad-hoc networks, Cellular networks, or Wireless Local Area Networks (WLANs) are illustrated elaborate.
- Several classical paradigms of GNNs applied in wireless networks are introduced to acquire a better understanding of the concepts and structures of GNNs.
- A comprehensive review of GNNs applied in wireless networks is summarized in terms of the existing directions, e.g., resource allocation and several emerging fields.
- Some challenges and potential research directions are summarized and discussed for the application of GNNs.

| Section III. Paradigms of GNNs | Section IV. Application in Wireless Networks |
|--------------------------------|---------------------------------------------|
| **A. Graph Convolutional Neural Networks**<br>1) Spectral-based GCNs<br>2) Spatial-based GCNs | **A. Resource Allocation**<br>1) Power control<br>2) Link scheduling<br>3) Others<br>4) Brief discussion |
| **B. Graph Attention Networks**<br>C. Spatial-Temporal Graph Neural Networks | **B. Emerging Fields**<br>1) Channel estimation<br>2) Traffic prediction<br>3) Vehicle communication<br>4) Others<br>5) Brief discussion |
| **D. Other Hybrid Methods** | |

Fig. 1: Road map of the overview.
in wireless networks.

The rest of this paper is organized as follows: In Section II, we introduce several basic definitions of graph-structured data and summarize the construction methods of WCG for Mesh/Ad-hoc networks and Cellular networks/WALNs. In Section III, we review several classical paradigms of GNNs that applied in wireless networks. In Section IV, we introduce the application of GNNs in wireless networks. In Section V, we discuss a few valuable directions for the application of GNNs in wireless networks. Finally, we conclude this work in Section VI. For ease of reading, the notations commonly used in graphs are summarized in Table I.

### TABLE I: COMMONLY USED NOTATIONS

| $G$ | Graph |
|-----|-------|
| $N$ | The number of nodes |
| $V$ | The set of nodes |
| $v$ | A node $v \in V$ |
| $N(v)$ | The set of neighbor nodes of $v \in V$ |
| $E$ | The set of edges |
| $e_{ij}$ | An edge from node $v_i$ to node $v_j$ |
| $A$ | Adjacent matrix |
| $A^T$ | The transpose of matrix $A$ |
| $A_{ij}$ | The element of the i-th row and the j-th column of matrix $A$ |
| $D$ | The degree matrix of matrix $A$ |
| $D^{-1/2}$ | The element $1/\sqrt{d}$ of the matrix $D$ |
| $X$ | Feature matrix |
| $X_i$ | The i-th column of matrix $X$ |
| $d$ | The dimension of the feature vector |
| $L$ | The maximum depth of GNNs |

## II. DEFINITION AND CONSTRUCTION OF GRAPH

In this section, we would like to first describe several basic definitions of graph-structured data. Then, the construction methods of WCG for different wireless network scenarios, e.g., Mesh/Ad-hoc networks, Cellular networks, and WLANs, are illustrated in detail.

### A. Definition of Graph

Graph-structured data is a kind of non-Euclidean data and is commonly expressed as $G = (V, E)$, where $V$ and $E$ are the sets of nodes and edges, respectively. Let $v_i \in V$ be a node and $e_{ij} = (v_i, v_j) \in E$ be an edge from node $v_i$ to node $v_j$. The adjacency matrix of a graph is represented as $A$. If $e_{ij} \in E$, $A_{ij} = 1$, otherwise, $A_{ij} = 0$. The graph is undirected if $A$ is symmetric, otherwise, the graph is directed. The degree matrix $D$ of an undirected graph is a diagonal matrix, where $D_{ii} = |N(v_i)|$. The Laplacian matrix of an undirected graph is defined as $L = D - A$. The normalized Laplacian matrix is defined as $\overline{L} = I_N - D^{-1/2}AD^{-1/2}$. Note that the normalized Laplacian matrix $L$ is a real semi-positive definite matrix. Accordingly, it can be decomposed into $L = UAU^T$, where $U$ is the eigenvector matrix and $\Lambda$ is a diagonal matrix of $[\Lambda]_{ii} = \lambda_i$ with $\lambda_i$ being the eigenvalue. While for the directed graph, the in-degree and the out-degrees are defined as $D_{ii}^{(in)} = \sum_i A_{ij}$ and $D_{ii}^{(out)} = \sum_i A_{ji}$, respectively. The transition probability matrix $P$ of a given directed graph is defined as $P = (D^{(out)})^{-1}A$. Accordingly, a symmetric normalized Laplacian of directed graph is defined as $\overline{L} = I - \frac{1}{n} (\Phi \Phi^T \Phi^T + \Phi^T \Phi^T \Phi \Phi^T)$, where $\Phi$ is generated according to $P$ and perron vector $\Phi_1$. It is worth mentioning that, from the reviewing results, there is almost no work to deal with directed graph based on spectral domain, but based on spatial domain. In a graph, each node may have its own attribute feature. The feature matrix of a graph is defined as $X \in \mathbb{R}^{N \times d}$. If feature matrix $X$ changes over time, the graph is defined as a spatial-temporal graph.

### B. Construction of Wireless Communication Graph

The first thing of using GNNs is to transform a wireless network into a graph. In general, according to the specific goal of the research, the topology structure of wireless networks may be constructed into an undirected graph or directed graph. According to the types of communication links and communication devices in wireless networks, it can be further constructed into a homogeneous or heterogeneous graph. In the sequel, the construction methods of WCG for various wireless network scenarios, e.g., Mesh/Ad-hoc networks, Cellular networks, or WLANs, are illustrated elaborately.

#### 1) Mesh/Ad-hoc Networks

Suppose there are $N$ transceiver pairs in homogeneous Mesh/Ad-hoc networks. To build a WCG for this kind of wireless network, we view the i-th transceiver pair as the i-th node of WCG, the feature vector of the i-th node includes the direct channel state information (CSI) $h_{ij}$ and other environmental information, such as the weight $w_i$ of the i-th node. The edge between nodes $v_i$ and $v_j$ in WCG may be undirected or directed. The feature vector of the undirected edge includes the interference CSI $h_{ij}$ and $h_{ji}$. While the feature vectors of two directed edges between nodes $v_i$ and $v_j$ can include $h_{ij}$ and $h_{ji}$, respectively. Fig. 2 shows a construction method of WCG for homogeneous Mesh/Ad-hoc networks with 3 transceivers.

Fig. 2: Illustration of constructing WCG for homogeneous Mesh/Ad-hoc networks with 3 transceivers.

In heterogeneous Mesh/Ad-hoc networks, suppose there are $N$ types of communication links. We treat the i-th transceiver pair with link type $m$ as node $v_{im}$ in the WCG. The set of...
neighboring nodes with link type \( n \) of node \( v_{im} \) are represented as \( \mathcal{N}^{(n)}_{im} \). The feature vector of node \( v_{im} \) includes the CSI \( \mathbf{h}_{im} \), direct link, and other environmental information corresponding to the link of type \( m \). The feature vector of the edge between nodes \( v_{im} \) and \( v_{jn} \) should also be considered from the perspective of undirected and directed. The feature vector of the undirected edge includes the CSIs, i.e., \( \mathbf{h}_{im,jn} \) and \( \mathbf{h}_{jn,im} \) of interference links. While the feature vectors of the directed edges between nodes \( v_{im} \) and \( v_{jn} \) include the \( \mathbf{h}_{im,jn} \) and \( \mathbf{h}_{jn,im} \), respectively. 

Fig. 3 illustrates a construction method of WCG for heterogeneous Mesh/Ad-hoc networks with 2 link types, where \( \mathbf{v}_{jn} \) is the feature vector of node \( v_{jn} \), \( \mathbf{e}_{jn,im} \) is the feature vector of the edge between nodes \( v_{jn} \) and \( v_{im} \), and “Link \( i \)” indicates the \( i \)-th communication link with type \( m \).

2) Cellular Networks/WLANs

In general, Cellular networks/WLANs may consist of \( M \) (\( M \geq 1 \)) access points (APs) and \( N \) (\( N \geq 1 \)) user equipments (UEs). Considering a simple situation including only one AP, which allocates the resources, such as power control and user association, etc., to the UEs. We treat the \( i \)-th UE as a node of WCG should be built while ignoring AP. The feature vector of the \( i \)-th node includes the CSI \( \mathbf{h}_{i,j} \) and other environmental information. The feature vector of the edge between nodes \( v_{i} \) and \( v_{j} \) includes the CSI \( \mathbf{h}_{ji} \) and \( \mathbf{h}_{ij} \), etc., which can also be ignored due to all UEs share one AP. An illustration of constructing WCG for Cellular networks/WLANs with a single AP is shown in Fig. 4.

Fig. 4: Illustration of constructing WCG for Cellular networks/WLANs with a single AP.

We further consider a more complex situation including multiple APs in Cellular Networks/WLANs, in which one AP may serve multiple UEs, and one UE may also access multiple APs. We first consider the scenario that one UE just accesses one AP and one AP serves multiple UEs. As shown in Fig. 5(a), UEs 1-4 and UEs 5-7 communicate with AP1 and AP2, while AP2 and AP1 are interfered with UE4 and UE5, respectively. The WCG of this kind of wireless networks can be built in two manners, which are illustrated in Fig. 5(b) and Fig. 5(c), respectively. In Fig. 5(b), there are two types of nodes indicating APs and UEs, respectively. The information of position, channel configuration, and device type are considered to be the feature vector of nodes. The feature vector of the edge between nodes \( v_{i} \) and \( v_{j} \) includes the direct/interference CSI and other link information. In contrast to Fig. 5(b), the APs are ignored in Fig. 5(c), which includes only one type of nodes indicating UE. The feature vector of a node includes the position of UE, channel configuration and device type, etc. The direct/interference CSIs and other link information can be considered to be the feature vector of edge. In practice, we can choose the appropriate manner according to the specific research tasks.

In addition to the Cellular networks/WLANs scenarios aforementioned, there are also other scenarios, such as one AP may serve multiple UEs and one UE may access multiple APs, called Heterogeneous Ultra-Dense Network (HUDN) [31]. We treat the UEs and APs as two types of nodes, as shown in Fig. 6(a). The feature vector of a node indicating UEs includes the CSIs to every AP. However, the feature vector of a node indicating APs includes the CSIs to every UE. The edge of WCG should be built only exists between nodes \( v_{i} \) and \( v_{j} \) of UEs that connect to the first-order neighborhood. Similarly, the corresponding neighborhood of AP is shown in Fig. 6(c). The details of designing the GNN model using this WCG can be found in [31], which exploit feature information from first-order and second-order neighborhood.
Various spectral-based GCNs have been defined by changing \( \mathbf{g} \). For example, Bruna et al. proposed spectral CNN (SpectralCNN) in which \( \mathbf{g} \) is learnable parameters \([32]\). However, due to the existing of the eigen-decomposition of \( \mathbf{L} \). SpectralCNN faces several challenges, such as low computational efficiency of eigen-decomposition and the learned graph filters cannot be applied in a graph with different structure \([27]\). To overcome these shortcomings, Defferrard et al. proposed ChebNet via redefining the graph filter with Chebyshev polynomials \([33]\). By constraining the number of parameters, Kipf et al. further proposed a model named GCN, which has the ability to overcome the overfitting, to minimize the number of operations at each layer, i.e.,

\[
\mathbf{X}_G = \mathbf{W} \left( \mathbf{I}_N + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \right) \mathbf{X},
\]

where \( \mathbf{W} \) is a learnable weight matrix. In order to tackle the case in which \( \mathbf{I}_N + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \in (0, 2) \) may lead to gradient explosion, the authors further transform \( \mathbf{I}_N + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \) into \( \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \), where \( \mathbf{A} = \mathbf{A} + \mathbf{I}_N \) and \( \mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij} \). Compared to GCN, ChebNet has higher computational complexity, but it has stronger expression ability. ChebNet’s \( K \)-order convolution operator can cover \( K \) steps neighbor nodes of the central node, while GCN only covers the first-order neighbor nodes. However, the perception domain of graph convolution can be expanded by stacking multiple GCN layers, so the flexibility is relative high.

Y. Ma et al. proposed a directed graph convolution network based on directed Laplacian, which is defined as \([28]\)

\[
\mathbf{Z} = \frac{1}{2} \left( \mathbf{\hat{\Phi}}^{-\frac{1}{2}} \mathbf{P} \mathbf{\hat{\Phi}}^{-\frac{1}{2}} + \mathbf{\hat{\Phi}}^{-\frac{1}{2}} \mathbf{P}^T \mathbf{\hat{\Phi}}^{-\frac{1}{2}} \right) \mathbf{XW},
\]

where \( \mathbf{D}_i^{(out)} = \sum_j \mathbf{A}_{ij} \) and \( \mathbf{P} = \left( \mathbf{D}_i^{(out)} \right)^{-1} \mathbf{A} \). The directed graph filter is approximated by the first-order Chebyshev polynomials.

\textbf{Remark 1.} The application of spectral-based GCNs in wireless networks will be introduced in Section IV. The definition of graph filters for spectral-based GCNs usually combines the adjacency matrix \( \mathbf{A} \) of wireless network topology and the channel state information \( \mathbf{H} \), which can make full use of the complex wireless information. On the other hand, \( \mathbf{H} \) is expressed as the propagation relationship between nodes from the aspect of wireless network environment.

2) Spatial-based GCNs

Spatial-based graph convolution is similar to the image convolution. The two convolution operations all extract the neighbor information of a node to obtain a richer feature representation of the node or the pixel. The difference between image convolution and spatial-based graph convolution is that the nodes in a graph are unordered while the pixels in an image are irregular, and the number of neighbors of each pixel in an image is limited while the number of neighbors of each node in a graph is not sure. So, spatial-based graph convolution operation cannot use a fixed-size convolution kernel like the image convolution operation. Thus, the key of spatial-based GCNs is to define the convolution operation with different neighborhood numbers and keep local invariance.
The most widely used spatial-based GCNs in wireless networks are message passing neural network (MPNN) and diffusion-convolutional neural networks (DCNNs). MPNN was proposed in [42], which is a unified framework of spatial-based GCNs, and decomposes the spatial-based graph convolution into a message aggregation phase and a combination phase, i.e.,

\[
\begin{align*}
\mathbf{m}_{i}^{(t)} &= \sum_{v_j \in N(v_i)} \mathcal{M}^{(t)}(\mathbf{X}_i^{(t-1)}, \mathbf{X}_j^{(t-1)}, \mathbf{e}_{ij}), \\
\mathbf{X}_i^{(t)} &= \mathcal{U}^{(t)}(\mathbf{X}_i^{(t-1)}, \mathbf{m}_{i}^{(t)}),
\end{align*}
\]

(4a, 4b)

where \( \mathbf{e}_{ij} \) is the feature vector of the edge between nodes \( v_i \) and \( v_j \), \( \mathcal{M}^{(t)}(\cdot) \) and \( \mathcal{U}^{(t)}(\cdot) \) are the aggregation function and the combination function in the \( t \)-th iteration, respectively. \( \mathbf{m}_{i}^{(t)} \) is the message aggregated from node \( v_i \)'s neighbors and \( \mathbf{X}_i^{(t)} \) is the hidden state of node \( v_i \) in the \( t \)-th iteration. It is observed that the computational efficiency of MPNN decreases with the increase of the number of nodes.

W. L. Hamilton further proposed Graph SAmpLe and aggregate (GraphSAGE) model via fixing the number of neighbors for message passing to overcome the shortcomings of MPNN [43]. The graph convolution operation of GraphSAGE is implemented by

\[
\mathbf{X}_i^{(t)} = \sigma \left( \mathbf{W}_i^{(t)} g_i^{(t)} \left( \mathbf{X}_i^{(t-1)}, \left\{ \mathbf{X}_j^{(t-1)}, \forall j \in S_{N(v_i)} \right\} \right) \right),
\]

(5)

where \( \mathbf{W}_i^{(t)} \) and \( g_i^{(t)}(\cdot) \) are a learnable weight matrix and an aggregation function in the \( t \)-th layer, respectively. \( \sigma(\cdot) \) is a nonlinear activation function. \( S_{N(v_i)} \) is a random sample of the node \( v_i \)'s neighbors. The main difference between MPNN and GraphSAGE is that GraphSAGE randomly samples a fixed number of neighbors for each node, while MPNN utilizes all the neighbors of each node. In addition, the diffusion-convolution operation in the DCNNs model builds a potential representation by scanning the diffusion process of each node through the transition probability matrix, i.e.,

\[
\mathbf{Z}_i^{(t)} = \sigma \left( \mathbf{W}_i^{(t)} \odot \mathbf{P}^t \mathbf{X}_i \right),
\]

(6)

where \( \mathbf{Z}_i^{(t)} \) is the hidden state of node \( v_i \) at the \( t \)-th layer. In addition, multi-head attention mechanism is further introduced to improve the expression ability of the attention layer, that is, \( K \) independent attention mechanisms can be utilized and then the output are concatenated together, i.e.,

\[
\mathbf{X}_i^{(t)} = \left\| \sum_{k=1}^{K} \alpha^{k} \left( \mathbf{X}_i^{(t-1)}, \mathbf{X}_j^{(t-1)} \right) \mathbf{W}^{(t-1)} \mathbf{X}_j^{(t-1)} \right\|,
\]

(7)

Remark 2. The GCN models introduced above meet the graph structures with the same node type and edge type. However, in wireless networks, the types of communication devices and of communication mechanisms between devices may be diverse, that is, the corresponding WCG may be a heterogeneous. Therefore, the design of heterogeneous graph convolution is helpful to learn different types of information and is more suitable for the business needs of practical network scenarios. In Section III, several works on designing graph convolution for heterogeneous WCG are introduced. However, the existing methods are designed based on spatial-based GCNs. There are few works on designing heterogeneous graph convolution based on spectral-based GCNs.

B. Graph Attention Networks

One notes that in a graph, different neighbor nodes generally have different influences on the central node. This implies that one needs to distinguish the influences of nodes with a proper means during the design process of the learning model. Attention mechanism [47] has been regarded as an expressive means of information fusion by assigning weight to given information. In recent years, attention mechanism is also introduced into GNNs, such as, Velickovic et al. proposed an attention mechanism based GNN model, i.e., Graph Attention Network (GAT), by adaptively allocating weight to different neighbors in the aggregation operation [48], which is defined as follows

\[
\mathbf{X}_i^{(t)} = \sigma \left( \sum_{j \in N(v_i)} \alpha(\mathbf{X}_i^{(t-1)}, \mathbf{X}_j^{(t-1)}) \mathbf{W}^{(t-1)} \mathbf{X}_j^{(t-1)} \right),
\]

where \( \alpha(\cdot) \) is the attention mechanism, \( \mathbf{X}_i^{(t)} \) is the hidden state of node \( v_i \) at the \( t \)-th layer. In addition, multi-head attention mechanism is further introduced to improve the expression ability of the attention layer, that is, \( K \) independent attention mechanisms can be utilized and then the output are concatenated together, i.e.,

\[
\mathbf{X}_i^{(t)} = \left\| \sum_{k=1}^{K} \alpha^{k} \left( \mathbf{X}_i^{(t-1)}, \mathbf{X}_j^{(t-1)} \right) \mathbf{W}^{(t-1)} \mathbf{X}_j^{(t-1)} \right\|,
\]

(7)

where \( \| \) denotes the concatenation operation and \( \alpha^{k}(\cdot) \) is the \( k \)-th attention mechanism. The illustration of single-head attention mechanism and multi-head attention mechanism are shown in Fig. 8 where \( \rho_{ij}^{(k)} \) denotes the attention weight between node \( v_i \) and node \( v_j \) obtained by the \( k \)-th attention mechanism. Other GNNs models using attention mechanism could be found in [49]–[52].

Remark 3. The essence of the attention mechanism is to weigh the information transmitted to distinguish the importance of different types of information. In wireless networks, especially heterogeneous wireless networks, the environmental information may be diverse, such as communication equipment, communication links, etc., which may affect the problems in different ways. Therefore, it is unreasonable to treat different types of information equally in the design of the learning model. On the contrary, one or more attention mechanisms should be designed according to the relevant prior knowledge to distinguish the influences of different types of information. There are few applications of graph attention mechanism in wireless networks.
C. Spatial-Temporal Graph Neural Networks

Spatial-temporal GNNs (STGNNs) play an important role in dealing with graphs that have dynamic node inputs while connected nodes are interdependent. There are two categories of STGNNs from the perspective of capturing temporal dependency, i.e., RNN-based methods [53–57] and CNN-based methods [58–62].

C. Chen et al. utilized the residual recurrent GNN (Res-RGNN) to predict the traffic flow in traffic network [58]. Res-RGNN utilizes the spatial attributes to capture the spatial features with diffusion convolution, while using graph recurrent unit (GRU) to discover the temporal dependency for each node. Specifically, the implementation of RGNN unit at time \( t \) is

\[
\begin{align}
\mathbf{r}^{(t)} &= \sigma \left( \Theta_r G \left[ \mathbf{x}^{(t)}, \mathbf{e}^{(t)}, \mathbf{s}^{(t-1)} \right] + \mathbf{b}_r \right), \\
\mathbf{u}^{(t)} &= \sigma \left( \Theta_u G \left[ \mathbf{x}^{(t)}, \mathbf{e}^{(t)}, \mathbf{s}^{(t-1)} \right] + \mathbf{b}_u \right), \\
\mathbf{c}^{(t)} &= \tanh \left( \Theta_c G \left[ \mathbf{x}^{(t)}, \mathbf{e}^{(t)}, \left( \mathbf{r}^{(t)} \odot \mathbf{s}^{(t-1)} \right) \right] + \mathbf{b}_c \right), \\
\mathbf{s}^{(t)} &= \mathbf{u}^{(t)} \odot \mathbf{s}^{(t-1)} + \left( 1 - \mathbf{u}^{(t)} \right) \odot \mathbf{c}^{(t)}, \\
\mathbf{y}^{(t+1)} &= \mathbf{W}_o \mathbf{s}^{(t)},
\end{align}
\]

where \( \mathbf{x}^{(t)}, \mathbf{e}^{(t)} \) and \( \mathbf{s}^{(t)} \) denote the graph signal, external feature and the outputted hidden state at time \( t \), respectively. \( \mathbf{r}^{(t)} \) and \( \mathbf{u}^{(t)} \) represent the reset gate and update gate at time \( t \), respectively. \( \Theta_r, \Theta_u, \text{ and } \Theta_c \) are the learnable graph filters, \( \mathbf{W}_o \) is the learned weights of the output layer. \( \mathbf{y}^{(t+1)} \) denotes the output at time \( t + 1 \).

S. Guo et al. proposed an attention mechanism based spatial-temporal GNN (ASTGCN) to predict traffic flow in traffic network [59]. Specifically, ASTGCN utilizes the spectral-based GCNs, i.e., ChebNet [33], to capture the spatial dependency among different nodes in traffic network graph. Meanwhile, one dimension CNN is utilized to capture temporal dependency for each node in time series. The implementation details of capturing spatial and temporal dependencies are illustrated in Fig. 9.

Remark 4. With the development of communication technologies, the wireless network becomes more complex and huge with massive terminals. As a result, the resource management of wireless networks becomes more and more challenging. To improve spectral efficiency, prediction plays an important role in wireless networks. Applying STGNNs to traffic prediction has attracted extensive attention in both academic and industry, which contributes to the resource management of wireless networks. Of course, other directions not involved in this overview also need to consider the spatial-temporal dependencies during designing the GNNs for wireless networks.

D. Other Hybrid Methods

Inspired by the conventional auto-encoders, graph auto-encoders (GAEs) utilizing GNNs as encoders to learn low-dimensional latent representations (or embeddings) of nodes have been investigated for wireless networks. The goal of encoders in GAEs is to encode the structural information of nodes. While the decoder in GAEs aims at decoding the structural information about the graph from the learned latent representations [63]. The general overview of GAEs is shown in Fig. 10. Specifically, the encoder maps node \( v \) to a low-dimensional embedding vector \( z_v \), based on the node’s structural information, and the decoder extracts the information interested from the low-dimensional embedding vector. GAEs have been used in many fields by virtue of their concise encoder-decoder structure and efficient encoding ability [64–69]. Kipf et al. proposed the variational GAE using a GCN encoder and a simple inner product decoder, which aims at the link prediction in citation networks [64]. The encoder maps each node to a low-dimensional latent representation using GCN, then a network embedding \( Z \) can be obtained. The decoder computes the pair-wise distance given network embedding and applies a non-linear activation. Finally, the decoder outputs the reconstructed adjacency matrix.

In recent years, reinforcement learning (RL) has been gradually applied to graph-structured tasks, such as graph generation [70], [71], graph classification [72], and graph reasoning tasks [73–78], etc. J. You et al. proposed a graph convolutional policy network (GCPN) based on RL and GCNs to address the problem of non-differentiable objective functions and constraints [71]. Graph attention model (GAM) was proposed based on RL and random walks to solve the graph classification task [72]. The generation of random walks was modeled as a partially observable Markov decision process. The RL agent performs two actions at each time step, i.e., predicts the label of input graph and generates the rank vector using designed rank network. The reward is designed as \( J(\theta) = \mathbb{E}_{P(S_{1:T}|\theta)} \sum_{t=1}^{T} r_t \), where \( r_t = 1 \) if the GAM classified the graph correctly, otherwise, \( r_t = -1 \). \( S_t \) is the environment. W. Xiong et al. proposed a DeepPath model to find the most informative path between two target nodes with the goal of solving the knowledge graph reasoning task [73]. The action of RL agents is to predict the next node in the path at each step and output a reasoning path in the knowledge graph.
graph. The reward functions include the scoring criteria: global accuracy, path efficiency and path diversity.

**Remark 5.** The main advantage of GAEs is to mine the topological information in the graph, and then learn an effective low dimensional feature vector representation for each node or the whole graph. This feature vector representation can reflect the characteristic that can separate from other nodes or graphs to a certain extent. Although RL has been widely used in wireless networks, the application of RL in GNNs is still in its infancy. Generally speaking, the introduction of RL can enable GNNs to achieve approximate optimal performance without the prior information of the environment, and have independent exploration and optimal decision-making capabilities. Therefore, the introduction of RL into wireless communication technology has important practical significance.

**IV. APPLICATIONS IN WIRELESS NETWORKS**

In this section, we focus on introducing comprehensively the application of GNNs in wireless networks. As shown in Table II, the applications of GNNs in wireless networks mainly cover resource allocation and a few emerging fields. The commonly used algorithms are illustrated in Fig. 11 for each research direction in wireless networks.

**A. Resource Allocation**

Resource allocation is one of the key issues for wireless communication systems. Applying GNNs to study the problem of resource allocation mainly focuses on power control, link scheduling, channel allocation, and spectrum allocation, etc.

1) **Power control**

A large amount of works has studied the power control problem using traditional optimization methods and DNNs. Unfortunately, the traditional optimization methods face high computational complexity. On the other hand, with the expansion of the wireless network scale, the scalability and generalization of DNNs will become worse. Motivated by these observations, many researchers utilize the GNNs, which have the natural characteristics of solving the problem with graph-structured data, to investigate the power control problem in wireless networks.

To develop scalable methods to solve the power control problem in wireless networks, Y. Shen et al. proposed an interference graph convolutional neural network (IGCNet) based on MPNN for K-user interference channels [29]. In particular, the K-user interference channels are modeled as a complete WCG with node and edge labels, as shown in Fig. 2. The aggregation and combination rules of IGCNet are designed as follows

\[
\gamma_{j,i}^{(t)} = MLP_1 \left( h_{i,j}, h_{i,j}, w_{i,j}, \beta_{j}^{(t-1)} \right), \quad (9a)
\]

\[
\alpha^{(t)}_i = \text{CONCAT} \left( \text{MAX}_{j \in N(v_i)} \left( \gamma_{j,i}^{(t)} \right), \sum_{j \in N(v_i)} \gamma_{j,i}^{(t)} \right), \quad (9b)
\]

\[
\beta^{(t)}_i = MLP_2 \left( \alpha^{(t)}_i, h_{i,i}, \beta^{(t-1)}_i, w_{i} \right), \quad (9c)
\]

where \( \text{MAX} (\cdot) \) is to take the largest value in a set, MLP1 and MLP2 represent two different MLPs, CONCAT denotes the operation of vector concatenations. \( \gamma_{j,i}^{(t)} \) denotes the feature vector of the edge connecting node \( v_j \) and node \( v_i \) in the \( t \)-th iteration. \( \alpha^{(t)}_i \) is the aggregated information from the neighbor nodes to the central node \( v_i \), and \( \beta^{(t)}_i \) is the updated hidden representation of node \( v_i \) in the \( t \)-th iteration. The IGCNet is trained in an unsupervised manner to learn the optimal power control.

In addition, a family of neural networks, i.e., message passing graph neural networks (MPGNNs), is designed to solve the problem of radio resource management in wireless networks [29]. It demonstrates that MPGNNs satisfy the permutation equivariance property and have the ability to address the resource management problem of large-scale wireless networks while enjoying a high computational efficiency. To guarantee an effective implementation, this
The proposed model, i.e., the random device (D2D) wireless networks, Spectral-based GCNs was employed in [80]–[82]. The effectiveness of WCGCN is evaluated with respect to the network type Algorithm Scheme

| Area | Year | Application | Wireless network type | Algorithm | Scheme |
|------|------|-------------|-----------------------|-----------|--------|
| Resource Allocation | 2019 | Power control in ad-hoc wireless networks | Mesh/Ad-hoc | Spatial-based GCNs | Y. Shen et al. |
| | 2020 | Power allocation with a set of transmitter receiver pairs in a large scale wireless network | Mesh/Ad-hoc | Spectral-based GCNs | M. Eisen et al. |
| | 2020 | Power allocation in wireless networks | Ad-hoc/Cellular | Spectral-based GCNs | M. Eisen et al. |
| | 2020 | Power control in ad-hoc wireless networks | Mesh/Ad-hoc | Spectral-based GCNs | M. Eisen et al. |
| | 2020 | Power control in decentralized wireless networks | Mesh/Ad-hoc | Spectral-based GCNs | I. Nikolić et al. |
| | 2021 | Power control/beamforming in heterogeneous D2D networks | Mesh/Ad-hoc | Spatial-based GCNs | X. Zhang et al. |
| | 2021 | Power control in multi-cell cellular networks | Cellular | Spatial-based GCNs | J. Guo et al. |
| | 2020 | Joint user association and power allocation in heterogeneous ultra dense networks | Cellular | Spatial-based GCNs | X. Zhang et al. |
| | 2020 | Resource allocation in free space optical (FSO) fronthaul networks | Mesh/Ad-hoc | Spatial-based GCNs | Z. Gao et al. |
| | 2021 | Resource allocation problems under asynchronous wireless network setting | Mesh/Ad-hoc | Spatial-based GCNs | Z. Wang et al. |
| | 2021 | Resource allocation in wireless IoT networks | Mesh/Ad-hoc | Spatial-based GCNs | T. Chen et al. |
| | 2020 | Power allocation in a single-hop ad-hoc wireless network | Mesh/Ad-hoc | Spectral-based GCNs | A. Chowdhury et al. |
| | 2019 | Link scheduling in D2D networks | Mesh/Ad-hoc | Graph embedding | W. M. Lee et al. |
| | 2020 | Schedule transmission for wireless networks in a distributed manner | Mesh/Ad-hoc | Spatial-based GCNs | Z. Zhao et al. |
| | 2019 | Temporal link prediction in various network systems | Mesh/Ad-hoc | Spatial-based GCNs | K. Lei et al. |
| | 2021 | Joint link scheduling and beam selection in ultra-dense D2D mmWave communication networks | Mesh/Ad-hoc | Spatial-based GCNs | S. He et al. |
| | 2020 | Channel allocation for densely deployed WLANs | WLANs | DRL with Spectral-based GCN | K. Nakashima et al. |
| | 2020 | Spectrum allocation in vehicle-to-everything (V2X) networks | Mesh/Ad-hoc | Spatial-based GCNs with Multi-agent RL | Z. He et al. |
| | 2021 | AP selection for Cell-Free massive MIMO systems | Cellular | Spatial-based GCNs | V. Ranasinh et al. |
| | 2021 | Intelligent reflecting surface and beamforming | Cellular | Spatial-based GCNs | T. Jiang et al. |
| | 2020 | Massive MIMO detection in wireless communication | Mesh/Ad-hoc | Spatial-based GCNs | A. Scotti et al. |
| | 2020 | Channel tracking for the massive MIMO networks | Cellular | GANE | Y. Yao et al. |
| | 2020 | Cellular traffic prediction | Cellular | Spatial-based GCNs and CNN with GLU | S. Zhao et al. |
| | 2020 | Satellite traffic prediction | Mesh/Ad-hoc | Spatial-based GCNs with graph GRU | L. Yang et al. |
| | 2020 | Multiagent cooperative control for CAV networks | Mesh/Ad-hoc | Spatial-based GCNs with RL | J. Dong et al. |
| | 2020 | Active traffic management for CAV networks | Mesh/Ad-hoc | Spatial-based GCNs with RL | P. Y. J. Ha et al. |
| | 2020 | Efficient point cloud processing | Mesh/Ad-hoc | Dynamic GCNs | J. Shao et al. |
| | 2020 | Point cloud delivery | Mesh/Ad-hoc | GANE | T. Fujishita et al. |
| | 2021 | 3D object detection | Mesh/Ad-hoc | 3D GNN | C. S. Jeong et al. |
| | 2019 | Throughput maximization for UAV assisted ground networks | Mesh/Ad-hoc | Spectral-based GCNs | S. Lohani et al. |
| | 2021 | Wireless network localization | - | Spectral-based GCNs | W. Yan et al. |
| | 2021 | Performance prediction in Next-Generation WLANs | WLANs | Spatial-based GCNs | P. Soto et al. |
| | 2021 | Routing in small satellite networks | Mesh/Ad-hoc | Spectral-based GCNs | M. Liu et al. |
| | 2021 | Decentralized control in wireless communication systems | Mesh/Ad-hoc | Spectral-based GCNs | M. Lee et al. |

To solve the power allocation problem for device-to-device (D2D) wireless networks, Spectral-based GCNs was employed in [80]–[82]. The proposed model, i.e., the random edge graph neural networks (REGNN), performs the convolutions over a random graph formed by the fading interference patterns in wireless networks. The authors further presented an unsupervised model-free primal-dual learning algorithm to train the weights of the REGNN to overcome the difficulties incurred by the constrained objective function. Additionally, REGNN is utilized to solve the problem of power control in decentralized wireless networks [83]. To adapt the time-varying topologies, the first-order meta-learning is adopted to adapt the new network configurations with a few shots exploiting the data obtained from multiple topologies. The problem of downlink power control in wireless networks over
a single shared wireless medium is investigated and addressed by using spectral-based GCNs and primal-dual learning [84]. The main highlight of the works aforementioned is to solve the resource management problem under the complex constraints via the primal-dual learning method in homogeneous wireless networks.

Compared with the homogeneous wireless networks, it is more challenging to design the GNNs-based learning mechanism for the resource allocation problem in heterogeneous wireless networks. X. Zhang et al. focused on addressing the problem of power control or beamforming using MPNN in heterogeneous D2D networks [50]. This work considers a heterogeneous D2D network with two types of links, in which each kind of link holds different features, as depicted in Fig. 3. In particular, let \( r = (n, m) \) be the interference from link type \( n \) to link type \( m \), the update rules in relation \((n, m)\) is defined as follows

\[
\begin{aligned}
\mathbf{e}_{j,n,m}[t] &= \phi_{(n,m)}^e (\mathbf{v}_{j,n}[t-1], \mathbf{e}_{j,n,m}[0]), \\
\mathbf{v}_{i,m}^{(n)} &= \phi_{(n,m)}^v \left( \mathbf{v}_{i,m}[t-1], \max_{j \in \mathcal{N}_m^{(n)}} \mathbf{e}_{i,j,m}[t] \right),
\end{aligned}
\]

where \( \phi_e^e \) and \( \phi_v^v \) are an edge update function and a node update function of relation \( r \), respectively. The aggregation rules is given by

\[
\mathbf{v}_{i,m}[t] = \rho_{i,m}^{e+v} \left( \left\{ \mathbf{v}_{i,m}^{(n)}[t] \right\}_n \right) = \frac{1}{c_{i,m}} \sum_n \mathbf{v}_{i,m}^{(n)}[t],
\]

where \( c_{i,m} \) is the number of relations causing interference to link \( i,m \). \( \rho_{i,m}^{e+v} (\cdot) \) is the aggregation function of node to node with link type \( m \).

Similarly, J. Guo et al. considered the power control problem in multi-cellular networks [85]. Specifically, this work models the cellular networks as a heterogeneous graph, i.e., wireless interference graph, and then proposed a heterogeneous GNN (HetGNN) based on spatial-based GCNs, called PGNN, to learn the power control policy in multi-cellular networks. Inspired by the finding that the parameter sharing scheme determines the invariance or equivalence relationship, the optimal power control policy has a combination of different PI and PE properties that existing heterogeneous GNNs do not satisfy [115]. Additionally, X. Zhang et al. considered the joint user association and power control problem in HUDNs [31]. The HUDNs are also modeled as a heterogeneous graph, which is shown in Fig. 6. A heterogeneous GraphSAGE (HGSAGE) that extended from GraphSAGE [43], is used to extract the latent node representations. To embrace both the generalization of the learning algorithm and the higher performance of HUDNs, the learning process of HUDNs is divided into two phases. The first phase of HUDNs learns a representation with a tremendous generalized ability to suit any scenario with different user distributions in an off-line manner. The second phase of HUDNs is to finely tune the parameters of GNN online to further improve the performance for quasi-static user distribution.

Z. Gao et al. investigated the optimal power assignment and node selection based on the instantaneous channel state information of the links in free space optical (FSO) fronthaul networks [86]. Spectral-based GCNs are utilized to exploit the FSO network structure with small-scale training parameters. Then, a primal-dual learning algorithm is developed to train the GNN in a model-free manner. Z. Wang et al. addressed the asynchronous decentralized wireless resource allocation problem with a novel unsupervised learning approach [87]. Specifically, the interference patterns between transmitting devices are modeled as a graph to capture the asynchrony patterns via the activation of the edges on a highly granular time scale. A decentralized learning architecture, i.e., the aggregation graph neural networks (Agg-GNNs) is designed based on the graph representation of interference and asynchrony. T. Chen et al. proposed a spatial-based GCNs based framework to address the high complexity of the practical implementation of wireless internet of things (IoT) networks [89]. The effectiveness of the framework is evaluated by the link scheduling in D2D networks and the joint channel and power allocation in D2D underlaid cellular networks.

The methods proposed by the aforementioned works all are based on data-driven neural networks with poor interpretability and scalability. Inspired by the algorithmic unfolding of the iterative WMMSE, i.e., unfolded WMMSE (UWMMSE), A. Chowdhury et al. proposed a data- and model-driven neural architecture to solve the power allocation problem in a single-hop Ad-hoc wireless network [90]. The optimization problem that should be solved is

\[
\begin{aligned}
&\min_{w, a, b} \sum_{i=1}^M (w_i q_i - \log w_i), \\
&\text{s.t.} \quad q_i = (1 - a_i h_i b_i)^2 + \sigma^2 a_i^2 + \sum_{i \neq j} a_i^2 h_{i j}^2 b_{i j}^2, \\
&\quad b_i^2 \leq p_{\text{max}},
\end{aligned}
\]

where \( w = [w_1, w_2, ..., w_M]^T, a = [a_1, a_2, ..., a_M]^T, b = [b_1, b_2, ..., b_M]^T \) are vectors of optimization variables. The allocated power is computed by a function \( p = \Phi (H; \theta_\phi, \theta_v) \) of the channel state matrix through a layered architecture \( \Phi \) with trainable weights \( \theta_\phi \) and \( \theta_v \). Precisely, setting \( b^{(t)} = \sqrt{p_{\text{max}}} \), the \( t \)-th layer of UWMMSE is implemented as follows

\[
\begin{aligned}
\bar{q}^{(t)} &= \Psi \left( H; \theta_\phi^{(t)} \right), \\
\nu^{(t)} &= \Psi \left( H; \theta_v^{(t)} \right), \\
\alpha^{(t)} &= \frac{h_{i i}^{(t-1)}}{\sigma^2 + \sum_j h_{i j}^2 b_{i j}^{(t-1)} b_{j j}^{(t-1)}}, \\
\bar{w}^{(t)} &= \frac{\bar{q}^{(t)}}{1 - \alpha^{(t)} h_i b_i^{(t-1)}} + \nu^{(t)}, \\
\bar{b}^{(t)} &= \alpha \left( \frac{a_i h_{i i} w_i^{(t)}}{\sum_j h_{i j}^2 a_j^{(t)} w_j^{(t)}} \right),
\end{aligned}
\]

and the output power is determined as \( p = \Phi (H; \theta_\phi, \theta_v) = (b^{(L)})^2 \). \( \alpha(z) := \frac{z}{\sqrt{z}} \) simply ensures that \( b_i^{(t)} \in [0, \sqrt{p_{\text{max}}}] \). The function \( \Psi \) parameterized by \( \theta_\phi \) and \( \theta_v \) is chosen to be spectral-based GCNs. The whole workflow of UWMMSE is shown in Fig. 12 which has better interpretability and scalability compared with the data-driven learning
models. Numerical experiments demonstrate that UWMMSE not only significantly reduced the computational complexity, but also improved the performance compared to the conventional WMMSE \[91\].

2) Link scheduling

Although the overall performance of wireless networks can be improved via power control, it is not enough to eliminate the strong interference for ultra-dense wireless networks. Link scheduling is regarded as an effective means to further improve the performance of wireless networks.

To overcome the high computational complexity of the traditional optimization methods and eliminate the costly channel estimation, M. Lee et al. proposed a novel DL-based graph embedding method to implement the link scheduling in D2D networks [92]. In detail, this work firstly models the D2D network as a fully connected directed graph, then computes a low-dimensional feature vector based on the distances of both communication and interference links without requiring the accurate channel state information for each node. Finally, a multi-layer classifier is utilized to learn the scheduling policy in a supervised and unsupervised manner, respectively. Numerical results show that this method can achieve near-optimal performance compared with the state-of-the-art methods but with a small number of training samples, and has competitive generalization and scalability.

A distributed scheduling scheme was proposed to overcome the difficulty encountered in solving the maximum weighted independent set (MWIS) problem for wireless networks [93]. The authors proposed a distributed MWIS solver based on spectral-based GCNs for link scheduling by combining the learning capabilities of GCNs and the efficiency of greedy MWIS solvers. The proposed solver achieves superior performance over greedy baselines with minimum increase in complexity, and generalizes well across different types of graphs and utility distributions.

Besides, the information of the dynamics, the topology structure and evolutionary patterns of dynamic networks can be fully exploited to improve the temporal link prediction performance. In dynamic network scenarios, K. Lei et al. introduced a novel non-linear GCN-GAN model by leveraging the benefits of spectral-based GCNs, LSTM as well as the GANs to tackle the challenging temporal link prediction task [94]. While in ultra-dense D2D mmWave communication networks, in order to effectively control the interference between communication pairs, spatial-based GCNs and primal-dual learning are utilized to solve the problem of joint beam selection and link activation across a set of communication pairs via inactivating part communication pairs [95].

3) Others

To improve the spectral efficiency in densely deployed WLANs, K. Nakashima et al. proposed a deep RL model using spectral-based GCNs for channel allocation [96]. The idea behind their work is that the objective function is modeled as a parametric function of topologies, channels and communication quality. Z. He et al. studied the spectrum allocation via learning the low dimensional representations of a graph by modeling the V2X network as a graph, where each vehicle-to-vehicle link is a node in the graph [97]. According to the learned characteristics, multi-agent RL is used for spectrum allocation. DQN is used to learn to optimize the total capacity of the V2X network. In [98], GraphSAGE is employed to predict the potential links between APs for cell-free massive MIMO.

4) Brief discussion

The application of GNNs in resource allocation mainly focuses on power allocation and link scheduling, while there is less work related to spectrum allocation and channel allocation. In addition, almost all works adopt GCNs and a few of them introduce RL in terms of the paradigms of GNNs. Although these works have achieved good results, there are also some issues needed to be considered. For one thing, resource allocation tasks in some wireless networks, such as intelligent factories and intelligent transportation, etc., may have many simple or complex constraints needed to be handled. Most of the existing works directly use projection strategy for simple constraints, while Lagrange dual learning framework is used for complex constraints. Although Lagrange dual learning framework can deal with complex constraints, it can not guarantee the complete satisfaction of constraints, and the training efficiency is not ideal. For another thing, in some wireless network scenarios with delay-sensitive traffics, the designed model should have low time complexity on the basis of achieving certain performance. A small amount of works has discussed the processing delay of the designed model, but there is still a certain distance from practical application. Therefore, the problem of model complexity is a direction worthy of exploration and research. On the other hand, most of the work is to solve the optimization problem of a single network layer. With the development of information and communication technologies, joint resource allocation at different network layers, i.e., cross-layer optimization, is a potential direction for the design of learning methods based on GNNs. The advantage of cross-layer optimization is that it can comprehensively utilize the information between different network layers, and then it may get performance improvement potentially.

B. Emerging Fields

GNNs are also applied in other wireless networks scenarios. Although these studies are in their infancy, the results of
the existing works show that GNNs have a good application prospect in these directions.

1) Channel estimation

Accurate CSI is important for guaranteeing the performance of massive MIMO high-dynamic networks. However, traditional solutions rely so much on hypothetical statistical models that they are hard to adapt the high-dynamic network environment. To overcome this defect, many researchers use DL to estimate channel of wireless communication system in recent years, but the ability of DL to extract spatial dependency is limited. However, GNNs have advantages in spatial dependency mining, so GNNs have been applied in this field in recent years.

The estimation of channels between the intelligent reflecting surface (IRS), the base station (BS), and the users is necessary for the optimal tuning of phase shifters at the IRS. J. Tang et al. proposed a DL model to configure the IRS and beamforming at the BS such that the system utility function is maximized directly based on the received pilots instead of the channel coefficients [99]. Specifically, spatial-based GCNs is utilized to directly map the received pilots to the beamformers at the BS and the reflective pattern at the IRS. While in [100], GAT is employed to solve the channel estimation for the two-way backhaul link of high-altitude platform stations with reconfigurable intelligent surfaces. Numerical results also show that even if the training data does not include all changes, the GAT estimator is robust to hardware impairments and small-scale fading characteristics changes. As a further case of channel control, A. Scotti et al. considered the inference task of massive MIMO detection under time-varying channels and higher-order quadrature amplitude modulation and proposed a message-passing solution based on GNNs, i.e., MIMO-GNN [101].

Y. Yan et al. proposed a new channel tracking method based on GAE [102]. Specifically, the channel tracking framework is designed as

\[
\hat{e}_{i,j}(t) = \text{MLP}_{e}^{\text{coder}}(\tilde{e}_{i,j}(t), \tilde{v}_i(t), \tilde{v}_j(t)), \quad (14a)
\]

\[
\hat{v}_i(t) = \text{MLP}_{v}^{\text{coder}}\left(\bar{v}_i(t), \sum_{\tilde{e}_{i,j}(t) \in \mathcal{N}(v_i)} \tilde{e}_{i,j}(t)\right), \quad (14b)
\]

\[
\tilde{e}_{i,j}(t) = \text{MLP}_{e}^{\text{decoder}}(\hat{e}_{i,j}(t)), \quad (14c)
\]

\[
\tilde{v}_i(t) = \text{MLP}_{v}^{\text{decoder}}(\hat{v}_i(t)), \quad (14d)
\]

where \(\tilde{v}_i(t)\) and \(\tilde{e}_{i,j}(n)\) are the resultant node feature of node \(i\) and edge feature between nodes \(i\) and \(j\) at time \(t\), respectively. MLP\(_{e}^{\text{coder}}\) and MLP\(_{v}^{\text{coder}}\) are coders for node and edge, respectively. In contrary, MLP\(_{e}^{\text{decoder}}\) and MLP\(_{v}^{\text{decoder}}\) are decoders. It’s not hard to find that the designed framework combines MPNN and codec. Numerical results confirm that the GNN-based scheme outperforms the feed-forward neural network in terms of the MSE.

2) Traffic prediction

Generally speaking, effective resource management can improve the utilization of network resources. In addition, if one can predict the required resource of future wireless traffic, resource management will become more flexible in wireless networks. However, the high spatial-temporal interdependencies make traffic prediction more challenging. There are fewer works on wireless network traffic prediction using GNNs, mainly including cellular network traffic prediction and satellite network traffic prediction.

To improve the accuracy of cellular traffic prediction, S. Zhao et al. proposed a new Spatio-Temporal GCNs incorporating Handover information (STGCN-HO) prediction model using the transition probability matrix of the handover graph [103]. STGCN-HO builds a stacked residual neural network structure that combines spectral-based GCNs and CNN with gated linear units [116] to capture the spatial and temporal interdependencies of traffic. Compared with RNN, STGCN-HO has a faster training speed due to the use of CNN, and has the ability to train or predict cell or base stations with the information collected from the entire graph at the same time. In addition, compared with CNN grid, STGCN-HO can predict both base stations and the cells within the base stations. While in reference to the satellite network traffic prediction, L. Yang et al. pointed out that the traditional network traffic prediction model could not effectively extract the spatio-temporal characteristics of network traffic. Therefore, they proposed a network traffic prediction model GCN-GRU via combining the spectral-based GCNs with GRU [104]. Specifically, GCN-GRU model utilizes the spectral-based GCN to extract the spatial characteristics of the satellite network traffic, and utilizes GRU model to extract the temporal characteristics of the satellite network traffic, and finally predict satellite network traffic through the fully connected layer. The illustration of GCN-GRU is shown in Fig. 13 where \(x(t), u(t), v(t)\) and \(s(t)\) correspond to (8a)-(8d) and \(X_t\) is the input feature at time \(t\).

3) Vehicle communication

Recently, GNNs have been applied to control the connected autonomous vehicles (CAVs) lane changing decisions for a road segment, to mitigate the highway bottleneck congestion, and to allocate spectrum in V2X networks. J. Dong et al. proposed a DL model that combines spectral-based GCN and a deep Q network to control multiple CAVs to make cooperative lane change decisions [105]. The graph construction methods of CAVs is shown in Fig. 14. There is a state \(S(t)\) that is con-
considered as a triplet at time $t$, i.e., $S^{(t)} = \{X^{(t)}, A^{(t)}, M^{(t)}\}$, where $X^{(t)}$, $A^{(t)}$ and $M^{(t)}$ denote the node feature matrix, adjacent matrix and a mask matrix that document the index of autonomous vehicles at time $t$, respectively. From the perspective of CAV operations, the proposed model not only enables CAV to successfully carry out lane changes to meet its personal intention of merging from the prescribed ramp, but also guarantees safety and efficiency. Similarly, RL algorithms are employed to train CAV driving behaviors, which can be used to relieve highway bottleneck congestion.  

4) Others

In addition to the aforementioned related works, GNNs are also used to solve other problems in wireless networks. J. Shao et al. proposed Branchy-GNN using the branch network and source-channel coding to reduce the computational cost and intermediate feature transmission overhead for efficient point cloud processing [107]. T. Fujihashi et al. proposed a novel soft point cloud transmission scheme that combines GNN-based point cloud coding and near-analog modulation for future wireless streaming of holographic and three-dimensional data [108]. C. S. Jeong et al. provided AR services via three-dimensional GNN using cameras and sensors on mobile devices [109]. S. Lohani et al. designed a model based on spectral-based GCNs to maximize the throughput of unmanned aerial vehicle (UAV)-assisted ground networks [110]. Spectral-based GCNs were utilized to solve the network localization problem of a wireless network in two-dimensional space [111]. In [112], spatial-based GCNs was first used to predict the achieved throughput in highly dense WLANs using channel bonding. Using the topology extraction ability of GNN, spectral-based GCNs based learning routing scheme was proposed to implement onboard routing in small satellite networks [113]. M. Lee et al. analyzed and enhanced the robustness of the decentralized GNN in different wireless communication systems, making the prediction results not only accurate but also robust to transmission errors [114].

5) Brief discussion

Emerging fields where GNNs are applied mainly include channel estimation, channel tracking, MIMO detection, traffic prediction, vehicle communication, point clouds, and so on. There is not much work in these fields, but some fields are worthy of further exploration. For instance, the traffic prediction task not only plays a pivotal role in the management of wireless network resources but also has higher requirements for data collection and acquisition. On the one hand, traffic prediction generally requires long-term historical data. It also involves issues such as the granularity of data collection, the level of data collection, and data privacy, which in turn brings varying degrees of difficulties. On the other hand, for scenarios with high requirements for traffic prediction service delay, the model designed is required to be as low as possible in complexity, so as to achieve real-time prediction capabilities. In addition, in terms of the application of the paradigms of GNNs in emerging fields, GCNs are frequently utilized. While there are also a small number of new paradigms of GNNs applied, such as GAT, GAE, and several generalized GCNs. According to the characteristics of different business needs, we can measure and compare different GNNs paradigms, and then adopt the best GNNs paradigm to better solve the problems faced.

V. Key Issues and Future Development

Although GNNs have made some progress in the application of wireless networks, some key issues need to be further studied in-depth. Accordingly, some ongoing or future research directions that are worth exploring are summarized as follows:

1) Acquisition of high-quality data

One of the fundamental elements of the data-driven DL method is to obtain a lot of training and testing data of wireless networks. The higher the quality of data, the better the training of the model. Though many researchers have studied the application of DL in wireless networks, most of the datasets used in the existing work are generated by numerical simulation, which is somewhat different from the real data of wireless networks. Furthermore, unlike the successful application of DL in image processing and social networks, etc, there is a lack of publicly recognized data set for the physical layer and media access layer of wireless networks. Therefore, it is urgent to collect and construct opening wireless network datasets for method comparison and performance verification.

Wireless networks have some unique characteristics, such as high dynamic, heterogeneous terminals, and non-uniform, resulting in many difficulties in obtaining the data of wireless networks, especially the real-time communication data of the physical layer and media access control layer. The data of wireless networks can generally be obtained from spectrum measurement instruments, base station, core network equipment, user terminals, and so on. Different data acquisition devices may be provided by different manufacturers, who may define different data extraction formats, feature names, and data calculation methods, and may also be different in the time granularity of extraction. This makes it very challenging to collect massive data in the real wireless network, especially at the physical layer and media access control layer. Meanwhile, the construction of opening testing and training datasets is still a very urgent and challenging task for the successful application of DL and GNN, etc, in wireless networks.

2) Distributed GNN learning model

As we all know, the goal of whether the data-driven DL methods or the data- and model-driven methods is to learn the super parameters of NNs that depend on the supercomputing power. Furthermore, the learning abilities of the DL methods are proportional to the network complexity. However, in wireless networks, the computing capacity of communication nodes is very limited, especially for the battery-powered lightweight devices. How to design a proper DL
model is a challenging and opening problem for the battery-powered lightweight devices. Fortunately, distributed ML is regarded as an effective and efficient technology to balance the performance and the computational resource and to reduce the required amount of training and testing data. However, several issues are needed to be considered when designing a GNN-based distributed learning model, such as the split of the GNN-based learning model, the parameter updating strategy, the integration of the results of each distributed running node, etc. In addition, the convergence of the whole distributed learning model is also should be guaranteed.

3) **Data privacy issues**

Generally speaking, the acquisition of the data of wireless networks inevitably touches the user’s privacy. However, privacy protection is one of the core issues in the field of information and communications technology in the future, and the process of data acquisition is cumbersome. Consequently, the acquisition of real network data will face many difficulties.

On the other hand, distributed learning model needs data interaction, which also involves data privacy. In some scenarios with high requirements for data privacy protection, data interaction security needs to be ensured. Some methods can be used to solve data privacy issues, such as secure multi-party computing and trusted execution environments, etc. Another popular solution is to use the federated learning, which has the outstanding advantage of privacy protection.

4) **Robustness of the learning model**

Although DL methods have been successfully applied in many fields, their effectiveness depends on high-quality training datasets. When the training dataset contains significant complex noise, anomaly intrusion, category imbalance, etc., its effectiveness cannot be guaranteed. This implies that the robustness of the learning model should be considered when designing the DL model for wireless networks. Generally, the robustness of the model can be discussed from two aspects, i.e., malicious or abnormal data and the defense ability of the model. On the one hand, when a malicious attacker attacks the model with input data, it can generally be defended by adversarial training, input transformation, gradient shielding, and detection and rejection. On the other hand, the model’s ability to resist attacks can be improved by improving the model itself and combining it with other security technologies.

5) **Scalability of the learning model**

In wireless networks, network scale, network state, and network service data volume may be highly dynamic, which requires stronger scalability of the designed model. For example, the changes in network scale (such as changes in the number of communication devices) and network state (such as changes in the location and attributes of network communication nodes) are enough to bring great challenges to the design of the learning model. Fortunately, GNNs have the ability to address these problems based on the existing research work, but the current work is still in its infancy and needs further research. On the other hand, with the high dynamic change of network business data volume, such as the sharp increase of business data volume, the real-time processing capability of the model should be enhanced accordingly.

6) **Cross-layer optimization**

Most of the existing works have covered different individual network layers in terms of GNNs applications. These algorithms generally cannot obtain the global optimal solution from the perspective of whole communication networks. More recently, with the development of information and communication technologies, joint resource allocation at different network layers, i.e., cross-layer optimization, is regarded as a potential direction for further improving the performance of learning methods. The advantage of cross-layer optimization is that the information at multiple network layers can be fully explored and exploited to design learning methods. Although cross-layer optimization has the potential to improve performance, it introduces more optimization parameters. Meanwhile, the objective function will be more complicated, which in turn requires higher computing power and may introduce relatively large processing delays. Parameter abstraction can reduce the complexity to a certain extent, but it may reduce the optimality of the generated configuration.

7) **Research of GNN-based deep unfolding method**

One shortcoming of the data-driven learning method is the lack of interpretability due to the NN-based methods treat the learning of mapping between the input and output as a black box, which weakens the domain knowledge. To tackle the shortcoming, an emerging direction is the algorithm unfolding (unrolling) aiming at combining the knowledge of data and domain fields. Motivated by the successful application of algorithm unfolding solving some classical problems like UWMMS, recently, deep unfolding is regarded as an effective combination to not only effectively utilize the interpretability and scalability of model-driven algorithms, but also use the expressive power of data-driven methods. Although deep unfolding has been paid a lot of attention from both the industry and academia, the problem to be solved has only a few simple constraints. Meanwhile, there is a little work to deeply study the GNN-based unfolding models to solve the optimization problem in the wireless network under practical constraints. This implies that there are still a lot of works to be explored for wireless networks.

8) **Model design with complex constraints**

So far, a large amount of work focuses on designing data-driven or data- and model-driven DL models aiming to solve the optimization problem of wireless network, however, these problem investigated is generally formulated subjecting to simpler constraints. In practical wireless communication, the transmission schemes may be investigated under some complex constraints, such as the per-antenna or per-base station power constraint, the quality of service constraint, delivery latency, etc. How to design an effective and efficient DL model under complex constraints is still an opening and challenging issue. Although Lagrange dual learning framework has been used to deal with such problems, there are still some problems with this method. On the one hand, the convergence speed and the effectiveness of the Lagrange dual learning framework are closely related to the updating step-sizes of the Lagrange multipliers and the model parameters, while the update step-sizes of the Lagrange multipliers and the model parameters are coupled with each other, which makes it difficult to
determine the approximate update step-sizes to maximize the performance of the DL model. On the other hand, the DL models trained based on Lagrange dual learning framework are generally difficult to fully satisfy constraints (especially for the problems with large-scale complex constraints), and the generalization performance of models is also questionable.

VI. CONCLUSIONS

In this paper, we first illustrated the construction method of WCG for various wireless networks. Then, we simply introduced several classical paradigms of GNNs that have been applied in wireless networks, and made a classified introduction for the application of GNNs in wireless networks, mainly including resource allocation and several emerging fields. From the overview results, the application of GNNs in wireless networks is still in its infancy. Many challenging problems are needed to be further solved and improved. Finally, based on the existing results, several key issues and research directions are summarized for participants interested in this domain.

REFERENCES

[1] S. He, et al., “A survey of millimeter-wave communication: Physical-layer technology specifications and enabling transmission technologies,” Proc. of the IEEE, vol. 109, no. 10, pp. 1666–1705, Oct. 2021.
[2] Q. Mao, et al., “Deep learning for intelligent wireless networks: A comprehensive survey,” IEEE Commun. Surveys Tuts., vol. 20, no. 4, pp. 2595–2621, Jul. 2018.
[3] C. Zhang, et al., “Deep learning in mobile and wireless networking: A survey,” IEEE Commun. Surveys Tuts., vol. 21, no. 3, pp. 2224–2287, Mar. 2019.
[4] H. Sun, et al., “Learning to optimize: Training deep neural networks for interference management,” IEEE Trans. Signal Process., vol. 66, no. 20, pp. 5438–5453, Aug. 2018.
[5] A. Bora, et al., “Compressed sensing using generative models,” in Proc. of the Inter. Conf. on Machine Learn. (ICML), Mar. 2017, pp. 537–546.
[6] N. Shlezinger, et al., “ViterbiNet: A deep learning based viterbi algorithm for symbol detection,” IEEE Trans. on Wireless Commun., vol. 19, no. 5, pp. 3319–3331, Feb. 2020.
[7] H. Kim, et al., “Communication algorithms via deep learning,” in Inter. Conf. on Learn. Represent., Feb. 2018.
[8] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in Proc. of the 27-th Inter. Conf. on Machine Learn. (ICML), Jul. 2010, pp. 399–406.
[9] M. Kulin, et al., “A survey on machine learning-based performance improvement of wireless networks: PHY, MAC and network layer,” Electronics, vol. 10, no. 3, pp. 1–63, Mar. 2021.
[10] F. Tang, et al., “Survey on machine learning for intelligent end-to-end communication towards 6G: From network access, routing to traffic control and streaming adoption,” IEEE Commun. Surveys Tuts., vol. 23, no. 3, pp. 1578–1598, Mar. 2021.
[11] F. Liang, et al., “Towards optimal power control via ensembling deep neural networks,” IEEE Trans. Commun., vol. 68, no. 3, pp. 1760–1776, Dec. 2020.
[12] W. Lee, et al., “Deep power control: Transmit power control scheme based on convolutional neural network,” IEEE Commun. Lett., vol. 22, no. 6, pp. 1276–1279, Apr. 2018.
[13] S. Zhang, B. Yin, and Y. Cheng, “Topology aware deep learning for wireless network optimization,” arXiv preprint arXiv:1912.08386, Jan. 2021.
[14] M. M.Bronstein, et al., “Geometric deep learning: Going beyond euclidean data,” IEEE Signal Process. Mag., vol. 34, no. 4, pp. 18–42, Jul. 2017.
[15] Z. Wu, et al., “A comprehensive survey on graph neural networks,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 1, pp. 4–24, Mar. 2021.
[16] Z. Zhang, et al., “Deep learning on graphs: A survey,” to be published in IEEE Trans. Knowl. Data En., Mar. 2020.
[17] F. Xia, et al., “Graph learning: A survey,” IEEE Trans. on Artifi. Intelli., vol. 2, no. 2, pp. 109–127, 2021.
[18] Z. Chen, et al., “Bridging the gap between spatial and spectral domains: A survey on graph neural networks,” arXiv preprint arXiv:2002.11867, Feb. 2020.
[19] R. Sato, “A survey on the expressive power of graph neural networks,” arXiv preprint arXiv:2003.04078, Mar. 2020.
[20] L. Ruiz, et al., “Graph neural networks: Architectures, stability, and transferability,” Proc. of the IEEE, vol. 109, no. 5, pp. 660–682, May 2021.
[21] S. Abadal, et al., “Computing graph neural networks: A survey from algorithms to accelerators,” arXiv preprint arXiv:2010.00130, Sept. 2020.
[22] Y. Yuan, et al., “Explainability in graph neural networks: A taxonomic survey,” arXiv preprint arXiv:2012.15445, Dec. 2020.
[23] J. Ye, et al., “How to build a graph-based deep learning architecture in traffic domain: A survey,” to be published in IEEE Trans. Intell. Transp. Syst., Dec. 2020.
[24] W. Liao, et al., “A review of graph neural networks and their applications in power systems,” arXiv preprint arXiv:2101.10025, Jan. 2021.
[25] Y. Gao, et al., “Deep learning on knowledge graph for recommender system: A survey,” arXiv preprint arXiv:2004.01387, Mar. 2020.
[26] L. Lamb, et al., “Graph neural networks meet neural-symbolic computing: A survey and perspective,” arXiv preprint arXiv:2003.00330, Feb. 2020.
[27] Z. Wu, et al., “A comprehensive survey on graph neural networks,” arXiv preprint arXiv:1901.00596, Jan. 2019.
[28] Y. Ma, et al., “Spectral-based graph convolutional network for directed graphs,” arXiv preprint arXiv:1907.08990, Jul. 2019.
[29] Y. Shen, et al., “A graph neural network approach for scalable wireless power control,” in Proc. of the IEEE Globecom Workshops (GC Wkshps), Dec. 2019, pp. 1–6.
[30] X. Zhang, et al., “Scalable power control/beamforming in heterogeneous wireless networks with graph neural networks,” arXiv preprint arXiv:2104.05463, Apr. 2021.
[31] J. He, et al., “Joint user association and power allocation in heterogeneous ultra dense network via semi-supervised representation learning,” arXiv preprint arXiv:2103.15367, Mar. 2021.
[32] J. Bruna, et al., “Spectral networks and deep locally connected networks on graphs,” in Proc. of the Inter. Conf. on Learn. Represent. (ICLR), Jan. 2014.
[33] M. Defferrard, et al., “Convolutional neural networks on graphs with fast localized spectral filtering,” in Proc. of the Inter. Conf. on Neural Inform. Process. Syst. (NIPS), Jun. 2016, pp. 3844–3852.
[34] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, Oct. 2016.
[35] M. Henaff, et al., “Deep convolutional networks on graph-structured data,” arXiv preprint arXiv:1506.05153, Jun. 2015.
[36] R. Li, et al., “Adaptive graph convolutional neural networks,” arXiv preprint arXiv:1801.00906, Jan. 2018.
[37] R. Levine, et al., “CayleyNets: Graph convolutional neural networks with complex rational spectral filters,” IEEE Trans. Signal Process., vol. 67, no. 1, pp. 97–109, Nov. 2018.
[38] S. Zhang, et al., “From spectrum wavelet to vertex propagation: Graph convolutional networks based on taylor approximation,” arXiv preprint arXiv:2007.00730, Nov. 2020.
[39] S. Fu, et al., “HesGCN: Hessian graph convolutional networks for semi-supervised classification,” Info. Sciences, vol. 514, no. 4, pp. 484–498, Apr. 2020.
[40] M. Xu, et al., “Spectral graph convolutional networks with lifting-based adaptive graph wavelets,” arXiv preprint arXiv:2108.01660, Aug. 2021.
[41] D. Duvenaud, et al., “Convolutional networks on graphs for learning molecular fingerprints,” arXiv preprint arXiv:1509.09292, Sept. 2015.
[42] J. Gilmer, et al., “Neural message passing for quantum chemistry,” in Proc. of the Inter. Conf. on Machine Learn. (ICML), Apr. 2017, pp. 1263–1272.
[43] W. L. Hamilton, et al., “Inductive representation learning on large graphs,” in Proc. of the Conf. and Workshop on Neural Inform. Process. Syst. (NIPS), Jun. 2017, pp. 1025–1035.
[44] J. Atwood, et al., “Diffusion-convolutional neural networks,” in Proc. of the Neural Inform. Process. Syst. (NIPS), Nov. 2016, pp. 1993–2001.
[45] H. Pei, et al., “Geom-GCN: Geometric graph convolutional networks,” in Inter. Conf. on Learn. Represent., Sept. 2019.
[46] T. Daniel, et al., “Spatial graph convolutional networks,” in Inter. Conf. on Neural Inform. Process., Nov. 2020, pp. 668–675.
et al.
P. Velickovic, et al.
X. Wang, et al.
C. Chen, et al.
S. Fang, et al.

B. Yu, H. Yin, and Z. Zhu, “Spatio-temporal graph convolutional network for traffic prediction,” in Proc. of the AAAI Conf. on Artif. Intell. (AAAI), vol. 33, no. 1, Jul. 2019, pp. 485–492.

X. Geng, et al., “Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting,” in Proc. of the AAAI Conf. on Artif. Intell., vol. 33, no. 01, Jan. 2019, pp. 3656–3663.

B. Yu, H. Yin, and Z. Zhu, “ST-Unet: A spatio-temporal u-network for graph-structured time series modeling,” arXiv preprint arXiv:1903.05631, Jul. 2019.

K. Guo, et al., “Optimized graph convolution recurrent neural network for traffic prediction,” IEEE Tran. on Intell. Transp. Syst., vol. 22, no. 2, pp. 1138–1149, Feb. 2020.

Y. Li, et al., “Diffusion convolutional recurrent neural network: Data-driven traffic forecasting,” in ICML, Feb. 2018.

B. Yu, H. Yin, and Z. Zhu, “Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting,” arXiv preprint arXiv:1709.04875, Jul. 2017.

S. Guo, et al., “Attention based spatial-temporal graph convolutional networks for traffic flow forecasting,” in Proc. of the AAAI Conf. on Arti. Intell. (AAAI), vol. 33, no. 1, Jul. 2019, pp. 922–929.

L. Ge, et al., “Temporal graph convolutional networks for traffic speed prediction considering external factors,” in 20th IEEE Int. Conf. on Mobile Data Manag. (MDM), Jun. 2019, pp. 234–242.

S. Fang, et al., “GSTNet: Global spatial-temporal network for traffic flow prediction,” in IJCAI, Aug. 2019, pp. 2286–2293.

Z. Wu, et al., “Graph wavenet for deep spatiotemporal graph modeling,” arXiv preprint arXiv:1906.00121, May 2019.

W. L. Hamilton, et al., “Representation learning on graphs: Methods and applications,” arXiv preprint arXiv:1709.05784, Sept. 2017.

T. N. Kipf and M. Welling, “Variational graph auto-encoders,” arXiv preprint arXiv:1611.07308, Nov. 2016.

S. Fan, et al., “Adversarially regularized graph autoencoder for graph embedding,” arXiv preprint arXiv:1802.04407, Feb. 2018.

J. J. Choong, X. Liu, and T. Murata, “Optimizing variational graph autoencoder for community detection,” in IEEE Int. Conf. on Big Data (Big Data), Dec. 2019, pp. 5353–5358.

T. H. Do, et al., “Graph auto-encoder for graph signal denoising,” in IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP), May 2020, pp. 3322–3326.

S. Fan, et al., “One2multi graph autoencoder for multi-view graph clustering,” in Proc. of The Web Conf. 2020, Apr. 2020, pp. 3070–3076.

I. Li, et al., “R-VGAE: Relational-variational graph autoencoder for unsupervised prerequisite chain learning,” in Proc. of the 28th Int. Conf. on Comput. Lang., Dec. 2020, pp. 1147–1157.

M. Svetlik, et al., “Automatic curriculum graph generation for reinforcement learning agents,” in Proc. of the AAAI Conf. on Artif. Intell., vol. 31, no. 1, Jan. 2017.

J. You, et al., “Graph convolutional policy network for goal-directed molecular graph generation,” in Proc. of the Int’l Conf. on Neural Info. Process. Syst. (NIPS), Jun. 2016, pp. 6412-6422.

J. B. Lee, et al., “Graph classification using structural attention,” in Proc. of the ACM SIGKDD Int’l Conf. on Knowl. Disco. & Data Min. (KDD), Jul. 2018, pp. 1666–1674.

W. Xiong, T. Hoang, and W. Y. Wang, “DeepPath: A reinforcement learning method for knowledge graph reasoning,” in Proc. of the Conf. on Empirical Methods in Natural Lang. Process. (EMNLP), Jul. 2017, pp. 557–563.

Q. Wang, Y. Hao, and J. Cao, “ADRL: An attention-based deep reinforcement learning framework for knowledge graph reasoning,” Knowledge-Based Systems, vol. 197, no. 6, p. 105910, Jun. 2020.

G. Han, et al., “Reasoning like human: Hierarchical reinforcement learning for knowledge graph reasoning,” in IJCAI, Jul. 2020, pp. 1926–1932.

P. Tiwari, H. Zhu, and H. Pandey, “DAPath: distance-aware knowledge graph reasoning based on deep reinforcement learning,” Neural Networks, vol. 135, no. 5, pp. 1–12, Mar. 2021.

S. Liu, et al., “MemoryNet: A deep reinforcement learning framework for incorporating memory component into knowledge graph reasoning,” Neurocomputing, vol. 419, no. 5, pp. 273–286, Jan. 2021.

L. Chen, et al., “RLPath: a knowledge graph link prediction method using reinforcement learning based attentive relation path searching and representation learning,” Applied Intell., pp. 1–12, Jul. 2021.

Y. Shen, et al., “Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis,” IEEE J. Sel. Areas Commun., vol. 39, no. 1, pp. 101–115, Nov. 2020.

M. Eisen and A. Ribeiro, “Large scale wireless power allocation with graph neural networks,” in Proc. of the IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC), Jul. 2019, pp. 1–5.

“Optimal wireless resource allocation with random edge graph neural networks,” IEEE Trans. Signal Process., vol. 68, no. 4, pp. 2977–2991, Apr. 2020.

“Transferable policies for large scale wireless networks with graph neural networks,” in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP), May 2020, pp. 3040–3044.

N. Nikolskaya, et al., “Power control adaptation via meta-learning for random edge graph neural networks,” arXiv preprint arXiv:2105.00459, May 2021.

N. Naderplashad, M. Eisen, and A. Ribeiro, “Wireless power control via counterfactual optimization of graph neural networks,” in Proc. of the IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC), Feb. 2020, pp. 1–5.

J. Guo and C. Yang, “Learning power control for cellular systems with heterogeneous graph neural networks,” in Proc. of the IEEE Wireless Commun. and Network. Conf. (WCNC), May 2021, pp. 1–6.

Z. Gao, M. Eisen, and A. Ribeiro, “Resource allocation via graph neural networks in free space optical fronthaul networks,” arXiv preprint arXiv:2106.15005, Jun. 2021.

Z. Wang, M. Eisen, and A. Ribeiro, “Unsupervised learning for asynchronous resource allocation in ad-hoc wireless networks,” in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP), Nov. 2021, pp. 8143–8147.

——, “Learning decentralized wireless resource allocations with graph neural networks,” arXiv preprint arXiv:2107.01489, Jul. 2021.

T. Chen, et al., “A GNN based supervised learning framework for resource allocation in wireless IoT networks,” to be published in IEEE Internet of Things Journal, Jun. 2021.

A. Chowdhury, et al., “Efficient power allocation using graph neural networks and deep algorithm unfolding,” in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Process (ICASSP), May 2021, pp. 4725–4729.

——, “Unfolding WMMSE using graph neural networks for efficient power allocation,” IEEE Trans. Wireless Commun., vol. 20, no. 9, pp. 6004–6017, Apr. 2021.

M. Lee, et al., “Graph embedding based wireless link scheduling with few training samples,” IEEE Trans. Wireless, vol. 20, no. 4, pp. 2282–2294, Dec. 2021.

Z. Zhao, et al., “Distributed scheduling using graph neural networks,” in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP), Nov. 2021, pp. 4720–4724.

K. Lei, et al., “GCN-GAN: A non-linear temporal link prediction model for weighted dynamic networks,” in Proc. of the IEEE Int. Conf. on Computer Commun. (INFOCOM), Jun. 2019, pp. 338–348, via meta-learning for random edge graph neural networks,” arXiv preprint arXiv:2105.00459, May 2021.

S. He, et al., “GBLinks: GNN-based beam selection and link activation for ultra-dense D2D mmWave networks,” arXiv preprint arXiv:2107.02412, Jul. 2021.

K. Nakashima, et al., “Deep reinforcement learning-based channel allocation for wireless lans with graph convolutional networks,” IEEE Internet Things Journal, vol. 8, no. 2, pp. 3182–3183, Feb. 2020.

Z. He, et al., “Resource allocation based on graph neural networks in vehicular communications,” in Proc. of the IEEE Global Commun. Conf. (Globecom), Dec. 2020, pp. 1–5.

V. Ranaasinghe, et al., “Graph neural network based access point selection for cell-free massive MIMO systems,” arXiv preprint arXiv:2107.02084, Jul. 2021.

T. Jiang, H. V. Cheng, and W. Yu, “Learning to reflect and to beamform for intelligent reflecting surface with implicit channel estimation,” IEEE J. Sel. Areas Commun., vol. 39, no. 7, pp. 1931–1945, May 2021.
[100] S. Lohani and R. T. Glasser, “Robust free space oam communications,” arXiv preprint arXiv:2010.12004, Oct. 2020.

[101] A. Scotti, et al., “Graph neural networks for massive MIMO detection,” arXiv preprint arXiv:2007.03703, Jul. 2020.

[102] Y. Yang, et al., “Graph neural network-based channel tracking for massive MIMO networks,” IEEE Commun. Lett., vol. 24, no. 8, pp. 1747–1751, Aug. 2020.

[103] S. Zhao, et al., “Cellular network traffic prediction incorporating handover: A graph convolutional approach,” in Proc. of the Annual IEEE Inter. Conf. on Sens., Commun., and Network. (SECON), Jun. 2020, pp. 1–9.

[104] L. Yang, et al., “A novel satellite network traffic prediction method based on GCN-GRU,” in Proc. of the Inter. Conf. on Wireless Commun. and Signal Process. (IC-WCSP), Oct. 2020, pp. 718–723.

[105] J. Dong, et al., “A DRL-based multiagent cooperative control framework for CAV networks: A graphic convolution Q network,” arXiv preprint arXiv:2010.05437, Oct. 2020.

[106] P. Y. J. Ha, et al., “Leveraging the capabilities of connected and autonomous vehicles and multi-agent reinforcement learning to mitigate highway bottleneck congestion,” arXiv preprint arXiv:2010.05436, Oct. 2020.

[107] J. Shao, et al., “Branchy-GNN: A device-edge co-inference framework for efficient point cloud processing,” in Proc. of the IEEE Inter. Conf. on Acoustics, Speech and Signal Process. (ICASSP), May 2021, pp. 8488–8492.

[108] T. Fujihashi, et al., “Wireless 3D point cloud delivery using deep graph neural networks,” arXiv preprint arXiv:2006.09835, Jun. 2020.

[109] C. S. Jeong, et al., “AR anchor system using mobile based 3D GNN detection,” J. Broadcast. Electron. Media, vol. 13, no. 1, pp. 54–60, Feb. 2021.

[110] S. Lohani and R. T. Glasser, “Robust free space oam communications with unsupervised machine learning,” in Proc. of the Frontiers in Optics (FIO), Sept. 2019, p. FTu5B.3.

[111] W. Yan, et al., “Graph neural network for large-scale network localization,” in Proc. of the IEEE Inter. Conf. on Acoustics, Speech and Signal Process. (ICASSP), Oct. 2021, pp. 5250–5254.

[112] P. Soto, et al., “ATARI: A graph convolutional neural network approach for performance prediction in next-generation WLANs,” Sensors, vol. 21, no. 13, p. 4321, May 2021.

[113] M. Liu, J. Li, and H. Lu, “Routing in small satellite networks: A GNN-based learning approach,” arXiv preprint arXiv:2108.08523, Aug. 2021.

[114] M. Lee, et al., “Decentralized inference with graph neural networks in wireless communication systems,” arXiv preprint arXiv:2104.09027, Apr. 2021.

[115] S. Ravanbaksh, et al., “Equivariance through parameter-sharing,” in Proc. of the Inter. Conf. on Machine Learn. (ICML), Feb. 2017, pp. 2892–2901.

[116] Y. N. Dauphin, et al., “Language modeling with gated convolutional networks,” in Proc. of the Inter. conf. on machine learn. (ICML), Dec. 2017, pp. 933–941.
Jiaheng Wang (M’10-SM’14) received the Ph.D. degree in electronic and computer engineering from the Hong Kong University of Science and Technology, Kowloon, Hong Kong, in 2010, and the B.E. and M.S. degrees from the Southeast University, Nanjing, China, in 2001 and 2006, respectively. He is currently a Full Professor at the National Mobile Communications Research Laboratory (NCRL), Southeast University, Nanjing, China. From 2010 to 2011, he was with the Signal Processing Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. He also held visiting positions at The Friedrich-alexander University of Erlangen-nurnberg, Germany, and the University of Macau, Macau.

His research interests are mainly on wireless communications and networks. Dr. Wang has published more than 100 articles on international journals and conferences. From 2014 to 2018, he served as an Associate Editor for the IEEE Signal Processing Letters. From 2018, he serves as a Senior Area Editor for the IEEE Signal Processing Letters. He is a recipient of the Humboldt Fellowship for Experienced Researchers.

Yongming Huang (M’10-SM’17) received the B.S. and M.S. degrees from Nanjing University, China, in 2000 and 2003, respectively. In 2007 he received the Ph.D. degree in electrical engineering from Southeast University, China. Since March 2007 he has been a faculty in the School of Information Science and Engineering, Southeast University, China, where he is currently a full professor. During 2008-2009, Dr. Huang was visiting the Signal Processing Lab, Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden. His current research interests include MIMO wireless communications, cooperative wireless communications and millimeter wave wireless communications. He has published over 200 peer-reviewed papers, hold over 60 invention patents. He submitted around 20 technical contributions to IEEE standards, and was awarded a certificate of appreciation for outstanding contribution to the development of IEEE standard 802.11aj. He has served as an Associate Editor for the IEEE Transactions on Signal Processing, and a Guest Editor for the IEEE Journal Selected Areas in Communications, is serving as an Editor-at-Large for the IEEE Open Journal of the Communications Society, and an Associate Editor for the IEEE Wireless Communications Letters.

Yaoxue Zhang (M’17-SM’18) received his B.Sc. degree from Northwest Institute of Telecommunication Engineering, China, in 1982, and his Ph.D. degree in computer networking from Tohoku University, Japan, in 1989. Currently, he is a professor with the School of Computer Science and Engineering, Central South University, China, and also a professor with the Department of Computer Science and Technology, Tsinghua University, China. His research interests include computer networking, operating systems, ubiquitous/pervasive computing, transparent computing, and big data. He has published over 200 technical papers in international journals and conferences, as well as 9 monographs and textbooks. Currently, he is serving as the Editor-in-Chief of Chinese Journal of Electronics. He is a fellow of the Chinese Academy of Engineering.