

Learning Wireless Power Allocation through Graph Convolutional Regression Networks over Riemannian Manifolds

Rashed Shelim, *Graduate Student Member, IEEE*, and Ahmed S. Ibrahim, *Member, IEEE*

Abstract—Optimum power allocation is a key enabler for maximizing data rate in wireless networks. Recently, various deep neural network models have been introduced for predicting power allocation in device-to-device (D2D) networks. However, they require large training samples (i.e., network layouts). On the contrary in this paper, we aim to develop a learning model for power allocation with fewer training samples, which is vital in dynamic networks (e.g., vehicular networks) with the need for fast learning of power allocation. The proposed model transforms Euclidean-based network layouts and power allocation problems into Riemannian (i.e., non-Euclidean) manifolds, which is shown to require fewer learning parameters and hence shorter learning time. Such transformation is possible thanks to the symmetric positive definite (SPD) property of spectral representation (i.e., Laplacian matrix) of network layouts. In particular, we propose a graph convolutional regression network (GCRN) for predicting power allocation over Riemannian manifolds in an unsupervised manner. Simulation results demonstrate that the proposed GCRN model approaches the maximum network rate in large-scale networks, with only 300 training samples as opposed to 10,000 in Euclidean-based learning models.

Index Terms—Device-to-device networks, graph convolutional networks, power allocation, Riemannian geometry, symmetric positive definite matrices.

I. INTRODUCTION

Power allocation in device-to-device (D2D) networks is one of the fundamental challenges in wireless communications, as it requires a judicious distribution of transmission powers to alleviate possible excessive interference among D2D pairs. With the goal of maximizing total network rate, the wireless power allocation problem can be defined as a non-convex optimization problem [1], [2]. Various solutions for the power allocation problem are based on optimization (e.g., [3]–[7]). Besides, modern machine learning (ML) based approaches, such as deep neural network (DNN) in [8], [9], or convolutional neural networks (CNN) in [10], have been considered for learning power allocation.

Graph neural networks (GNN) is of special importance for learning power allocation as it matches the natural graph

structure of wireless connectivity patterns [11]–[14]. For example in [15], a GNN architecture was introduced for learning power allocation in large-scale wireless networks. While such GNN-based learning model approaches the total network rate, it requires 10,000 training samples (i.e., network layouts). Dynamic networks (e.g., vehicular ones) change its topology and size (i.e., number of vehicles) frequently. Maximizing the total network rate with such dynamic variations requires fast learning of any new network topology with fewer training samples, and this is the *motivation* of this paper.

Euclidean-based wireless network layouts and power allocation problem can be transformed into Riemannian (i.e., non-Euclidean) manifolds, which is shown to require fewer learning parameters and hence shorter learning time. Structural properties (i.e., connectivity patterns between network nodes) of wireless networks can be effectively represented through regularized weighted Laplacian matrices [16], [18], [19]. The regularized weighted Laplacian matrices are Symmetric positive definite (SPD) one which can be represented as points over Riemannian manifolds [20], [21]. To this end, we propose to apply graph convolutional regression networks (GCRN) toward learning topological representations over Riemannian manifolds, and such learning model is used to predict power allocation in an unsupervised manner.

Riemannian geometry has recently been used to address challenges in wireless communication systems, such as the link scheduling to maximize sum rate performance in [16]–[18], the deployment of relays in [19] or the design of beamforming codebooks in [22], [23]. Nonetheless, no prior research work has considered the Riemannian geometry to reduce the number of training samples for power allocation problems.

In this paper, we first model local topology around each communication pair over the Riemannian manifold through regularized weighted Laplacian matrices, with weights that are proportional to its channel state information (CSI). Then we use a GCRN model that learns the local topological modeling over Riemannian manifolds in an unsupervised manner for predicting power allocation, with differentiable exponential and logarithmic maps, and this is the *novel* contribution of this paper. The proposed GCRN model enjoys two unique advantages for solving power allocation problems. First, it is stable to imperfect CSI. Second, it can be generalized to large-scale wireless networks. The proposed GCRN model can also be trained in a supervised learning manner using weighted minimum mean square error (WMMSE) [3] as the ground truth which is often considered as performance upper bound [15].

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work is supported in part by the National Science Foundation under award number CNS-2144297.

Rashed Shelim and Ahmed S. Ibrahim are with the Department of Electrical and Computer Engineering, Florida International University, Miami, FL-33174, USA

Corresponding author: Rashed Shelim (e-mail: rshel019@fiu.edu).

To provide a theoretical guarantee, we show that the proposed GCRN satisfies permutation equivariance property in wireless network graphs and thus makes it amenable to train and test in different network topologies independent of its shape and size. Finally, we show that the proposed method is competitive with the state-of-the-arts by using only 300 training samples.

In brief, the main contributions of this paper are as follows:

- We model the local topology around each communication pair through the regularized weighted graph Laplacian matrices and use the modelings in the proposed GCRN method to learn wireless power allocation over Riemannian manifolds in an unsupervised manner. The proposed GCRN model can also be trained in a supervised learning manner using WMMSE [3] as the ground truth.
- We show that the proposed GCRN satisfies the permutation equivariance property which allows to train and execute it across different network topologies.
- We verify our proposed method with numerical experiments in simulated D2D wireless networks scenarios. Simulated results demonstrate that the proposed method approaches the sum rate maximization performance as the state-of-the-arts with only 300 training samples.

The remainder of this paper is organized as follows. We present the system model and problem formulation in Section II. In Section III, we provide the details of the proposed GCRN architecture and analyze its permutation equivariance property. In Section IV, we present our simulation results. We draw the conclusion in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we first briefly describe the relevant background for the proposed method. We then present a system model of the D2D wireless networks on the Riemannian manifold and formulate the wireless power allocation problem.

A. Preliminaries

A differentiable manifold $\mathcal{M}$ is a topological space [24] where each point p has a neighborhood that is topologically equivalent to a Euclidean space. A manifold locally similar to Euclidean space refers to tangent space and its properties as vector space. In particular, the tangent space $T_p\mathcal{M}$ of a differential manifold $\mathcal{M}$, at any point p , is a vector space of all possible vectors v that are tangent to manifold $\mathcal{M}$ at point p . A Riemannian manifold $(\mathcal{M}, g)$ is a smooth and real manifold $\mathcal{M}$ equipped with a positive definite inner product g_p at each point p on the tangent space $T_p\mathcal{M}$ and is studied by Riemannian geometry [24]–[26]. This implies that we have a notion of distance that satisfies the metric properties within a locally Euclidean-like, although it is not exactly Euclidean space. The $n \times n$ SPD matrices $Sym_n^{++} = \{S \in \mathbb{R}^{n,n} | S = S^T, \text{all eigen values of } S \text{ are positive}\}$ forms a cone-like manifold which is special class of Riemannian manifold [25], [27]. The geodesic that connects two SPD points $S_i, S_j \in Sym_n^{++}$ can be represented by Riemannian metric such as log-Euclidean metric (LEM) [28] and is given by

$$d(S_i, S_j) = \|\log(S_i) - \log(S_j)\|_F^2 \quad (1)$$

where, $i \neq j$, and $\|\cdot\|_F$ denotes the Frobenius matrix norm.

Two important operations that connect the manifold $\mathcal{M}$ and the tangent space $T_p\mathcal{M}$ at a point $p \in \mathcal{M}$ are exponential and logarithmic maps for which SPD manifold is endowed analytical formula. The exponential map $\exp_p : T_p\mathcal{M} \rightarrow \mathcal{M}$ defines a unique geodesic distance (shortest curve) from a point p in the direction of a vector v in $\mathcal{M}$ which results in a point in $\mathcal{M}$ and is given by [29]

$$\exp_p(v) = p^{1/2} \text{Exp}(p^{-1/2} v p^{-1/2}) p^{1/2}. \quad (2)$$

The logarithm map $\log_p : \mathcal{M} \rightarrow T_p\mathcal{M}$ at point p is the inverse of the exponential map which corresponds to a point $y \in \mathcal{M}$ when sent back via exponential map and is given by [29]

$$\log_p(y) = p^{1/2} \text{Log}(p^{-1/2} y p^{-1/2}) p^{1/2}. \quad (3)$$

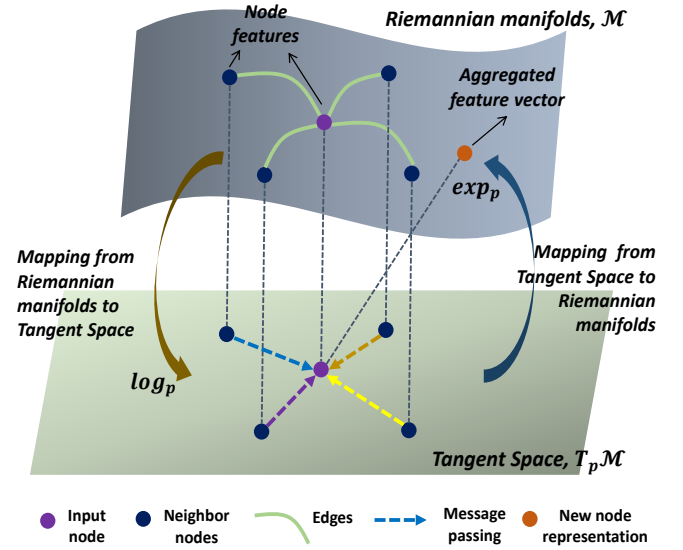


Fig. 1: An illustration of basic operations of a Riemannian graph convolutional layer. In each layer, node representations are computed by aggregating messages from their neighbor nodes.

Riemannian graph convolutional networks [30] generalizes the notion of graph convolutional networks of the form in [31] such that the network operates on Riemannian manifolds with differentiable exponential and logarithmic maps. To perform graph convolution, node representations are computed by aggregating messages from their neighbor nodes. As the tangent space of a point on Riemannian manifolds is always Euclidean, functions with trainable parameters are executed over Euclidean space. Let, $\mathcal{I}(u)$ be the set of neighbors of u and σ be the pointwise non-linear function. Then the propagation rule for each node $u \in V$ is calculated as

$$X_u^{(k+1)} = \sigma \left(\exp_p \left(\sum_{v \in \mathcal{I}(u)} \tilde{A}_{vu} \Theta^{(k)} \log_p(X_v^{(k)}) \right) \right), \quad (4)$$

where σ is the non-linear activation function, $\Theta^{(k)}$ is the learnable parameters, and $\tilde{A} = D^{-1/2}(A + I)D^{-1/2}$ is normalized adjacency matrix which captures the graph connectivity where $D \in \mathbb{R}^{n,n}$ denotes the diagonal degree matrix

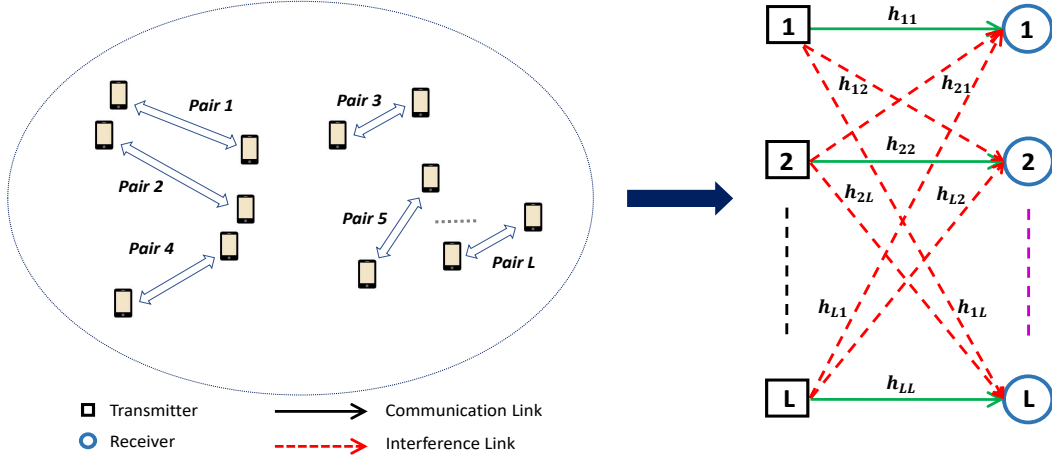


Fig. 2: A sample L -user interference channel and its corresponding graph.

with $D_{uu} = \sum_v (A_{vu} + I_{vu})$ and $D_{uv} = 0$ if $u \neq v$. At layer k , the propagature rule in (4) maps each node representation $\mathbf{X}_v^k \in \mathcal{M}$, where $v \in \mathcal{I}(u)$ is a neighbor of u , to the tangent space of a chosen point $p \in \mathcal{M}$ using the logarithmic map $\log_p$. Then an exponential map $\exp_p$ is applied to map the linearly transformed tangent vector back to the Riemannian manifold. The basic operations of the Riemannian graph convolutional layer are illustrated in Fig. 1.

Standard regression method, on the other hand, can be operated on Riemannian manifolds [30] by extending the underlying idea of radial basis function [32], [33] to Riemannian manifolds. More specifically, a list of centroids $\mathcal{C} = [c_1, c_2, \dots, c_{|\mathcal{C}|}]$, are first jointly learned with the Riemannian graph convolutional network using backpropagation, where each $c_m \in \mathcal{M}$. Then the pairwise distance between c_i and $\mathbf{X}_j^K$ (i.e., geodesic distance) is calculated as: $\psi_{ij} = d(c_i, \mathbf{X}_j^K)$. Then all distances $(\psi_{1j}, \psi_{2j}, \dots, \psi_{|\mathcal{C}|j}) \in \mathbb{R}^{|\mathcal{C}|}$ are concatenated to summarize the position of $\mathbf{X}_j^K$ with respect to the centroids. The node level regression can be defined as

$$\hat{y} = \Theta_0^T (\psi_{1j}, \psi_{2j}, \dots, \psi_{|\mathcal{C}|j}), \quad (5)$$

where $\Theta_0 \in \mathbb{R}^{|\mathcal{C}|}$ and T denotes matrix transposition.

B. Local topology modeling over Riemannian Manifolds

Fig. 2 illustrates an L -user interference channel, the underlying wireless networks model of L communication pairs. Each communication pair $\mathcal{D}_q$, where $q = 1, 2, \dots, L$, has a transmitter and a receiver represented in black square and blue circle, respectively. We represent the communication links between any transmitter-receiver pair with solid green lines. Whereas the interference links to and from neighbor communication pairs are represented with red dashed lines. With full frequency reuse, when each transmitter communicates with its paired receiver, it causes interference to its neighbor receivers.

The wireless networks of L communication pairs in Fig. 2 includes $n = 2L$ nodes, which are L transmitters and L receivers. From this model, local topology around each communication pair $\mathcal{D}_q$ can be modeled as a *weighted* and *undirected* graph $G_{\mathcal{D}_q}(V, E_{\mathcal{D}_q})$, as shown in Fig. 3, where

$V = \{v_1, v_2, \dots, v_n\}$ is the set of all n nodes and $E_{\mathcal{D}_q}$ is the set of all m_q edges (i.e., links) that are connecting the transmitter of $\mathcal{D}_q$ pair to its intended receiver along with the interference links to and from its neighbor communication pairs. The *weight* of edges is set to the instantaneous CSI $h_{i,j}$ between its communication pair. We employ ϵ -neighborhoods model [34], in which interference links between any transmitter v_i and receiver v_j of its neighbor communication pairs are considered if the distance $\|v_i - v_j\|^2 < \epsilon$, where $i \neq j$, and $i, j \in L$. The ϵ -neighborhoods-based local topology modeling is reasonable as the interference caused by any transmitter to its neighbor receiver is negligible if the transmitter is placed far from its neighbor receiver.

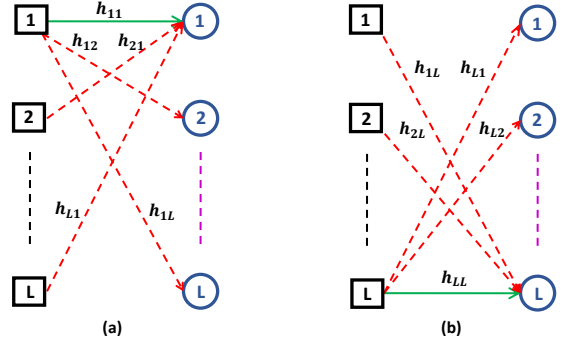


Fig. 3: Example of local topology modeling of: a) communication pair $\mathcal{D}_{q=1}$, and b) communication pair $\mathcal{D}_{q=L}$.

For an edge t , $1 \leq t \leq m_q$, connecting transmitter v_i and receiver v_j , where $\{v_i, v_j\} \in V$, we define an edge vector, where the i -th and j -th elements are given by $a_{ti} = 1, a_{tj} = 1$ and rest of the entries are zero. The incidence matrix $\mathbf{A}_{\mathcal{D}_q} \in \mathbb{R}^{n \times m_q}$ of graph $G_{\mathcal{D}_q}$ is the matrix with t -th column given by edge vector $\mathbf{a}_t$. The weight matrix $\mathbf{W}_{\mathcal{D}_q} \in \mathbb{R}^{m_q \times m_q}$ is defined as a diagonal matrix, whose diagonal entry (i.e., (t, t) -th element) is equal to the weight of the t -th edge. Finally, the graph Laplacian matrix $\mathbf{L} \in \mathbb{R}^{n \times n}$ can be computed as [16], [18]

$$\mathbf{L}_{\mathcal{D}_q} = \mathbf{A}_{\mathcal{D}_q} \mathbf{W}_{\mathcal{D}_q} \mathbf{A}_{\mathcal{D}_q}^T, \quad (6)$$

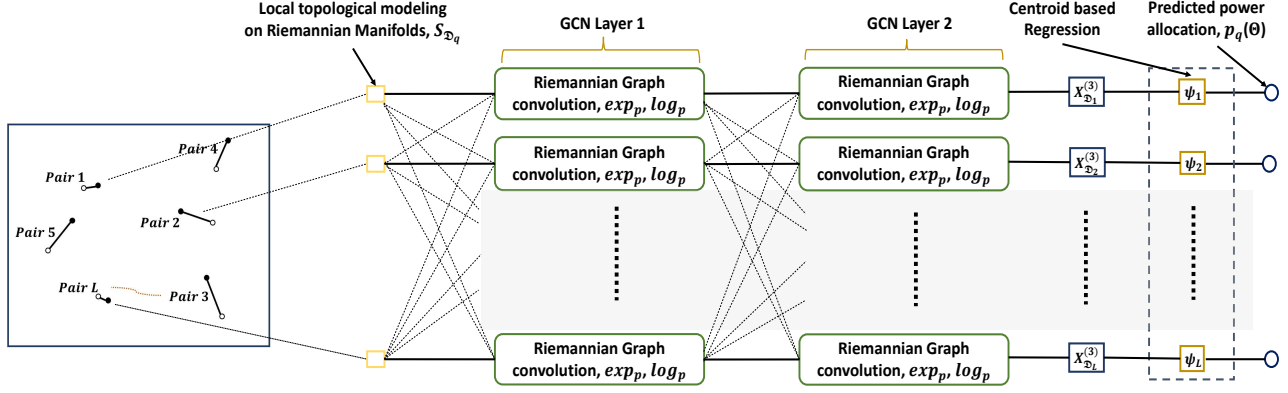


Fig. 4: Overall proposed GCRN model as applied to power allocation problem with two Riemannian GCN layers and a centroid based regression layer.

where $q = 1, 2, \dots, L$, and T denotes matrix transposition. The graph Laplacian matrices are positive semi-definite ones. With a simple regularization step by adding a scaled identity matrix to the $\mathbf{L}_{\mathcal{D}_q}$, we can produce a regularized graph Laplacian matrix [35] such as

$$\mathbf{S}_{\mathcal{D}_q} = \mathbf{A}_{\mathcal{D}_q} \mathbf{W}_{\mathcal{D}_q} \mathbf{A}_{\mathcal{D}_q}^T + \gamma \mathbf{I}_n, \quad (7)$$

where $\gamma > 0$ is a regularization parameter which is an arbitrarily small scalar and $\mathbf{I}_n \in \mathbb{R}^{n,n}$ is the identity matrix. The regularized graph Laplacian matrices $\mathbf{S}_{\mathcal{D}_q}$ are SPD one [20], [21], which are special class of Riemannian manifolds [25]. To perform graph convolution for power allocations, we consider each communication pair as a node in a graph. Hence, the node feature of each communication pair is given by the local topological modeling in (7).

C. Problem Formulation

We define the power allocation problem in the L -user interference channel, previously shown in Fig. 2, as the sum rate maximization problem of L communication pairs. Denote $\mathbf{p} = [p_1, p_2, \dots, p_L]$ as the power allocation vector. With full frequency reuse over bandwidth B , our goal is to find the optimal combinations of power allocation vector to maximize the summation of instantaneous information-theoretic rates and the problem is formulated as

$$\begin{aligned} \max_{\mathbf{p}} \quad & \sum_{q=1}^L B \log_2 \left(1 + \frac{p_q |h_{qq}|^2 \rho_{qq}^{-\alpha}}{\sum_{i \neq q} p_i |h_{iq}|^2 \rho_{iq}^{-\alpha} + \sigma_q^2} \right), \quad (8) \\ \text{s.t.} \quad & 0 \leq p_q \leq P_{\max}, \forall q = 1, 2, \dots, L, \end{aligned}$$

where p_q , h_{iq} , ρ_{iq} are the transmission power, fast-fading channel gain and Euclidean distance, respectively, between the q -th transmitter and i -th receiver. Moreover, α is the path loss exponent and σ_q^2 denotes the noise variance.

III. RIEMANNIAN GCN MODEL FOR POWER ALLOCATION

In this section, we describe the proposed wireless power allocation scheme for D2D wireless networks. We start by explaining the proposed GCRN method and then provide its key property.

A. Proposed GCRN method

To perform graph convolution over the Riemannian manifold for power allocation, we considered each communication pair as a node in the graph. For each communication pair, the node feature is extracted through local topological modeling over Riemannian manifold which is described in Section II. For instance, we first model the local topology around each communication pair from the L -user interference channel (Fig. 2) as a weighted and undirected graph (Fig. 3) which captures the amount of interference a communication pair imposes and receives to and from its neighbor communication pairs. Then with a simple regularization step, we transformed the model into Riemannian manifolds as a single point through regularized graph Laplacian matrix which is SPD one. Then we used these SPD points as the input to the Riemannian GCN for power allocation.

We use two Riemannian graph convolutional layers and a centroid based regression layer (as described in Section II) in our proposed GCRN architecture for predicting the power allocation as shown in Fig. 4. To perform graph convolution, we view the q -th communication pair as the q -th node in a graph. Since distant communication pairs cause little interference, we only consider an edge from v to q if the distance between transmitter v and receiver q is below a certain threshold ϵ . The node feature is given by the local topological modeling $\mathbf{S}_{\mathcal{D}_q}$, and the graph connectivity is captured by the normalized adjacency matrix $\tilde{\mathbf{A}} = \mathbf{D}^{-1/2}(\mathbf{A} + \mathbf{I})\mathbf{D}^{-1/2}$ [30], [31].

Let, $\mathbf{X}_{\mathcal{D}_q}^{(k)}$ denotes the representation of q -th node at the k -th graph convolutional layer, where $k = 0, 1, 2$. As initialization, the representation of q -th node is set by its local topological modeling $\mathbf{X}_{\mathcal{D}_q}^{(0)} = \mathbf{S}_{\mathcal{D}_q} \in \text{Sym}_n^{++}$. By doing so, we represent each node on Riemannian manifolds. We then obtain a new node representation of q at the next step by aggregating all messages (i.e., SPDs) from its neighbor nodes before applying the non-linear activation function σ . Let $\mathcal{I}(q)$ be the set of neighbors of q -th node. Then the propagation rule for each node $q \in L$ is calculated as [30]

$$\mathbf{X}_{\mathcal{D}_q}^{(k+1)} = \sigma \left(\exp_p \left(\sum_{v \in \mathcal{I}(q)} \tilde{\mathbf{A}}_{vq} \boldsymbol{\Theta}^{(k)} \log_p(\mathbf{X}_{\mathcal{D}_v}^k) \right) \right), \quad (9)$$

where $\Theta^{(k)} \in \mathbb{R}^{d,d}$ is the learnable weight matrix in k -th layer. Here, at layer k , the representation $\mathbf{X}_{\mathcal{D}_v}^k$ that lies in Riemannian manifolds, where v is a neighbor of q , is mapped to the tangent space of a chosen point $p \in \mathcal{M}$ using the logarithmic map $\log_p$ as in (3). We do this mapping because the tangent space of a point on Riemannian manifolds is Euclidean, so functions with trainable parameters are executed there. Then an exponential map $\exp_p$ is applied to map the linearly transformed tangent vector back to the Riemannian manifold as in (2). Furthermore, when applying the non-linear activation function directly on Riemannian manifold $\mathcal{M}$, we ensure that its application is manifold preserving, i.e., $\sigma : \mathcal{M} \rightarrow \mathcal{M}$. In particular, we use *ReLU* as the non-linear activation function since the output of *ReLU* is still an SPD matrix [36], [37].

The output of the Riemannian graph convolutional networks consists of a set of SPDs corresponding to L communication pairs $\{\mathbf{X}_{\mathcal{D}_1}^{(3)}, \mathbf{X}_{\mathcal{D}_2}^{(3)}, \dots, \mathbf{X}_{\mathcal{D}_L}^{(3)}\}$. We apply the centroid based regression method [30] over Riemannian manifold to predict the power allocation from these SPDs. More specifically, a list of centroids $\mathcal{C} = [c_1, c_2, \dots, c_{|C|}]$, are first jointly learned with the Riemannian graph convolutional network using backpropagation, where each $c_m \in \text{Sym}_n^{++}$. Then the pairwise distance between c_i and $\mathbf{X}_q^{(3)}$ (i.e., geodesic distance) is calculated using LEM [28]

$$\psi_{ij} = \|\log(c_i) - \log(\mathbf{X}_q^{(3)})\|_F^2, \quad (10)$$

where, $\|\cdot\|_F$ denotes the Frobenius matrix norm. Then all distances $(\psi_{1q}, \psi_{2q}, \dots, \psi_{|C|q}) \in \mathbb{R}^{|C|}$ are concatenated to summarize the position of $\mathbf{X}_q^{(3)}$ with respect to the centroids. Using this, the node level regression is applied to produce a continuous scalar output and is given by

$$p_q(\Theta) = (\Theta^{(3)})^T (\psi_{1q}, \psi_{2q}, \dots, \psi_{|C|q}), \quad (11)$$

where $\Theta^{(3)} \in \mathbb{R}^{|C|}$. The purpose of this regression layer is to embed the resulting node representation $\mathbf{X}_{\mathcal{D}_q}^{(3)}$ containing the node features that live on Riemannian manifolds into a scalar value presenting the predicted transmission power of the corresponding communication pair.

Finally, noting that maximizing the sum rate in (8) be the ultimate goal of the power allocation problem, we apply the following loss function at the last layer of the GCRN to train the model as in [9], [10], [15],

$$l(\Theta) = -\mathbb{E}_{\mathbf{h}} \left\{ \sum_{q=1}^L \log_2 \left(1 + \frac{|h_{qq}|^2 p_q(\Theta)}{\sum_{i \neq q} |h_{iq}|^2 p_i(\Theta) + \sigma_q^2} \right) \right\}, \quad (12)$$

where $p_q(\Theta)$ is the transmission power generated by the proposed model. Note that no labels are needed to train the model by adopting this loss function. Thus, the model is trained in an unsupervised manner.

The power allocation method by the proposed GCRN is summarized in Algorithm 1. As shown in the Algorithm, if variable *phase* is set to "train", then the algorithm calculates the loss and updates the learnable weights Θ to minimize the loss. Otherwise, the algorithm considers it the "test" phase; no loss function is applied. Hence, no weights are updated.

Algorithm 1 Pseudo-Code of Power Allocation by GCRN

```

1: Inputs: Data point  $\{\mathbf{S}_{\mathcal{D}_q}^y \in \text{Sym}_n^{++}, q \in L, y = 1, \dots, Y\}$ ,
2:           and phase;
3: Outputs: Predicted transmission power of  $L$  D2D pairs;
4: Process:
5: for  $y = 1, \dots, Y$  do
6:   for  $q = 1, \dots, L$  do
7:      $\mathbf{X}_{\mathcal{D}_q}^{(0)} \leftarrow \mathbf{S}_{\mathcal{D}_q}^y \in \text{Sym}_n^{++}$ ; ▷ Initialization
8:      $\mathbf{X}_{\mathcal{D}_q}^{(2)} \leftarrow \sigma \left( \exp_p \left( \sum_{v \in \mathcal{I}(q)} \tilde{\mathbf{A}}_{qv} \Theta^{(1)} \log_p(\mathbf{X}_{\mathcal{D}_v}^{(1)}) \right) \right)$ ; ▷ GCN
9:      $\mathbf{X}_{\mathcal{D}_q}^{(3)} \leftarrow \sigma \left( \exp_p \left( \sum_{v \in \mathcal{I}(q)} \tilde{\mathbf{A}}_{qv} \Theta^{(2)} \log_p(\mathbf{X}_{\mathcal{D}_v}^{(2)}) \right) \right)$ ; ▷ GCN
10:     $p_q(\Theta) \leftarrow (\Theta^{(3)})^T (\psi_{1j}, \psi_{2j}, \dots, \psi_{|C|j})$ ; ▷ Regression
11:  end for
12:  if phase="train" then
13:    Calculate loss  $l(\Theta)$  in (12) and update weights  $\Theta$ ;
14:  end if
15: end for
16: Return:  $\{p_q^y(\Theta)\}_{q=1}^L, y = 1, \dots, Y$ .
```

As mentioned earlier, the GCRN can also be trained in a supervised learning manner. Explicitly, we use WMMSE [3] as the ground truth which is often considered as performance upper bound [15]. Then we apply mean squared error (MSE) loss function (i.e., L2 norm) as the loss function, and is given by

$$l(p_{\text{WMMSE}}, p_q(\Theta)) = \frac{\sum_{q=1}^L (p_{\text{WMMSE}} - p_q(\Theta))^2}{L}, \quad (13)$$

where p_{WMMSE} is the transmit power of WMMSE scheme. However, such approach is suitable when the labeled training is available, which is computationally challenging especially when the number of nodes is large [9].

B. Permutation Equivariance Property of Proposed GCRN Model

The proposed GCRN has *permutation equivariance* property, with respect to the underlying graph structure of L -user interference channel. To conveniently describe these properties, we can rewrite the input-output relation of GCNs in (9) in a more compact form $\mathcal{X} = \Phi(L(\mathbf{A}), \Psi, \mathbf{X})$, where $L(\mathbf{A})$ is the normalized graph Laplacian matrix which captures the graph connectivity, Ψ is a tensor gathering the learnable weights $\Theta^{(i)}, i = 1, 2$ at all two graph convolutional layers, and $\mathbf{X}$ is the tensor gathering the of nodes features (i.e., SPDs) in which component $\mathbf{X}_q$ is associated with node q and are supported on graph through graph connectivity. In graph theory, functions that are equivariant to permutations imply permuting the input graph consistently permutes the output. Putting it from a wireless communication perspective, the optimal power allocation scheme requires to entail permutation invariant power allocation policy [14], [15] so that the allocated power is not affected by the indexing of the communication nodes. This is a key feature since the node indexing is arbitrary for different wireless network layouts. So, permutation equivariance is important for the proposed GCRN model as it makes training model not depending on indexing

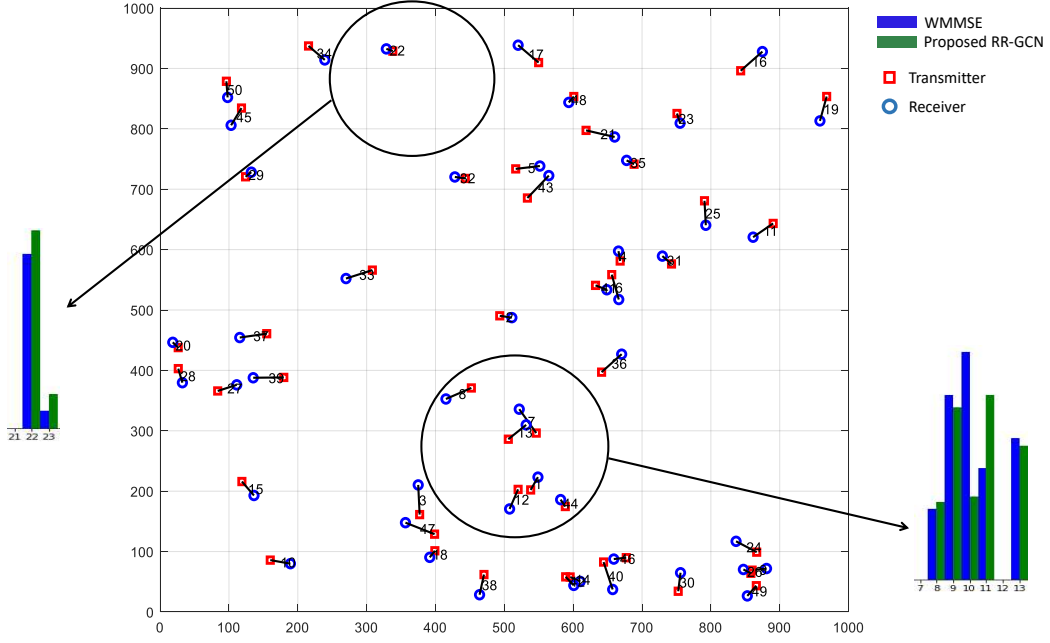


Fig. 5: Example of power allocation by proposed GCRN methods and WMMSE [3] for $L = 50$ pairs.

of the nodes. This property is hard-coded in the structure of the proposed GCRN, as shown by the following proposition:

Proposition 1. (*Permutation Equivariance in proposed GCRN:*) Consider regularized graph Laplacians $\mathbf{L}(\mathbf{A})$ and $\hat{\mathbf{L}}(\hat{\mathbf{A}})$ along with tensors of node features $\mathbf{X}$ and $\hat{\mathbf{X}}$ for which there exists a permutation matrix $\mathbf{\Pi} \in \mathcal{Q}_n$, where $\mathcal{Q}_n$ is the set of all permutation matrices, such that $\hat{\mathbf{A}} = \mathbf{\Pi} \cdot \mathbf{A} = \mathbf{\Pi}^T \mathbf{A} \mathbf{\Pi}$ and $\hat{\mathbf{X}} = \mathbf{\Pi} \cdot \mathbf{X} = \mathbf{\Pi}^T \mathbf{X} \mathbf{\Pi}$, where " $\mathbf{\Pi} \cdot$ " denotes permuted matrices over nodes, then we have

$$\hat{\Phi}(\hat{\mathbf{L}}(\hat{\mathbf{A}}), \Psi, \hat{\mathbf{X}}) = \mathbf{\Pi} \cdot \Phi(\mathbf{L}(\mathbf{A}), \Psi, \mathbf{X}). \quad (14)$$

Proof. The proof is immediate, by observing that $\mathbf{L}(\mathbf{\Pi} \cdot \mathbf{A}) = \mathbf{\Pi} \cdot \mathbf{L}(\mathbf{A})$ [38], thus $\hat{\Phi}(\hat{\mathbf{L}}(\hat{\mathbf{A}}), \Psi, \hat{\mathbf{X}}) = \Phi(\mathbf{\Pi} \cdot \mathbf{L}(\mathbf{A}), \Psi, \mathbf{\Pi} \cdot \mathbf{X}) = \mathbf{\Pi} \cdot \Phi(\mathbf{L}(\mathbf{A}), \Psi, \mathbf{X})$, and permutation commute with the pointwise non-linear activation function (i.e., ReLU) and regression layer. $\square$

The permutation equivariance property allows to train and execute the proposed scheme across different network topologies independent of its shape and size. Simulation results in Section IV.B show its effectiveness to generalize to large-scale networks.

IV. PERFORMANCE EVALUATION

In this section, we present simulation results of our proposed power allocation scheme for sum rate maximization in L -user interference channel. We provide simulation results for different topological scenarios and compare these results against state-of-the-arts.

A. Simulation Setup

We consider L -communication pairs within a square sized wireless network area. In particular, we randomly deploy

L transmitters on the $A \times A$ area, while the receivers are uniformly distributed around its corresponding transmitters within the range between $d_{\min}$ to $d_{\max}$. In the simulation, we consider a distance-based path-loss according to ITU-1411 outdoor model. Then we add the shadowing and Rayleigh fast fading into the channel model. We consider the WMMSE [3] as the performance upper bound (i.e., benchmark) where perfect CSI is assumed to be available to WMMSE. We randomly generate 100 wireless network layouts for testing, and all results presented in this section are averaged over 100 test wireless network layouts and normalized by the average sum rate of WMMSE [3]. The simulation parameters are summarized in Table I.

TABLE I: NETWORK SIMULATION PARAMETERS

Parameter	Value
Bandwidth, B	5MHz
Carrier Frequency	2.4 GHz
Transmitter and Receiver Antenna height	1.5m
Transmitter and Receiver Antenna Antenna gain	2.5 dB
Noise spectral density	-169 dBm/Hz

To train the proposed GCRN model, we use Adam optimizer [39] for 20 iterations where the initial learning rate is set to 0.001 and the weight decay rate is set to 0.0005 based on the experiments. Moreover, the dropout rate after the first GCRN layer is set to 0.1 for regularization. The training parameters of the proposed GCRN model are summarized in Table II.

B. Simulation Results

We first experiment with the impact of the number of training samples on the sum rate performance of the proposed method for 50 communication pairs. From Table III, the sum rate performance for 300 training samples is enough, provided that no significant performance gain is achieved by using more than 300 training samples.

TABLE II: REGRESSION GRAPH CONVOLUTIONAL NETWORKS PARAMETERS

Parameter	Value
Input Matrix size	Defined by the dimension of input SPDs
Number of iterations	20
Learning rate	0.001
Weight decay	0.0005
Dropout rate	0.1

TABLE III: IMPACT OF THE NUMBER OF TRAINING SAMPLES AS PERCENTAGE OF WMMSE [3]

Number of Training Network Layouts	160	250	300	500	700	1000
Average Sum Rate (%)	86.25	91.84	93.82	93.12	92.56	91.64

To show that the proposed method is capable of learning the state-of-the-art optimization strategy, we randomly generate a wireless network layout for $L = 50$ communication pairs in $1000m \times 1000m$ wireless network area as illustrated in Fig. 5 and compare the power allocation between our proposed method and benchmark WMMSE [3] algorithm. As shown in the figure, the proposed GCRN method produces a power allocation pattern close to the WMMSE output. This indicates that our proposed scheme can learn the WMMSE optimization strategy.

In Table IV, we summarize the sum rate performances and approaches of various power allocation schemes for large-scale networks, including GCN and deep neural network (DNN) based solutions for $L = 50$ communication pairs at the area of $1000m \times 1000m$ wireless network layout. From Table IV, the proposed GCRN can achieve 93.82% of average sum rate generated by WMMSE [3] by only using 300 training samples. It can outperform the DNN based PCNet [9], which requires 100,000,000 training samples, and All Active¹ algorithm by a large margin. Compared with the WCGCN [15] which requires 10,000 training samples to achieve 97.50% of the average sum rate generated by WMMSE [3], we use far fewer training samples but with only 3.68% of performance loss of sum rate. Moreover, it can be seen from the table that the proposed GCRN takes the least number of learning parameters (e.g., 946) as opposed to 1792 in WCGCN [15] to achieve comparable performance against the state-of-the-arts.

1) *Stability to Imperfect CSI*: To experiment the stability of proposed GCRN to imperfect CSI, we model the environmental noise by adding additive white Gaussian noise (AWGN) $\mathcal{A}_G \sim (0, \sigma^2)$ to the perfect CSI (i.e., instantaneous estimated CSI). We also experiment with the partial CSI (i.e., slowly varying large-scale channel parameters) where we only add the shadowing effect to ITU-1411 path-loss channel model and no small-scale fast-fading is considered. We measure the stability of the proposed method by four realizations for different values of AWGN standard deviation σ and partial CSI. We still consider 50 communication pairs in a $1000m \times 1000m$ region for each scenario and the results are summarized in Table V.

TABLE IV: POWER ALLOCATION SCHEMES WITH AVERAGE SUM RATES FOR 50 COMMUNICATION PAIRS

Method	CSI Used	Average Sum rate	Number of samples	Number of parameters	Approach
WMMSE [3]	Yes	100%	/	/	Iterative, Optimization
Proposed GCRN (Unsupervised)	Yes	93.82%	300	946	Graph Modeling, & GCNs
WCGCN [15] (Unsupervised)	Yes	97.50%	10,000	1792	Graph Modeling, & GCRN
PCNet [9] (Unsupervised)	Yes	79.70%	100,000,000	/	Multilayer perceptron
All Active	No	73.65%	/	/	All active links with $P_{\max}$ power

We can observe from the table that the sum rate performance slightly but gradually decreases with the increase of standard deviation values but still achieves good performances. The performance is also satisfactory for partial CSI.

TABLE V: AVERAGE SUM RATE PERFORMANCE WITH IMPERFECT CSI FOR 50 COMMUNICATION PAIRS

AWGN Standard Deviation	Perfect CSI	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$	Partial CSI
Average Sum Rate %	93.82%	92.42%	90.23%	89.16%	88.86%	87.76%

2) *Scalability to Different Number of Pairs and Pairwise Distances*: In this section, we experiment with the scalability of our proposed method to two important topological scenes: 1) how it performs when the number of communication pairs increases and 2) its performance when pairwise distance changes.

We first test the scalability of GCRN to topologies with different numbers of communication pairs in the network area of $1000m \times 1000m$. In this scalability testing, we keep the number of pairs the same in both testing and training datasets. As shown in Table VI, PCNet [9] achieves near-optimal sum rate performance when the network is small. As the network becomes large, the performance of PCNet [9] approaches to Strongest² which is a simple baseline algorithm. This shows that the PCNet method can hardly learn any critical interference information that exists in the networks. On the other hand, the performance of the proposed GCRN is stable as the network size increases. The proposed method even achieves higher performance than both PCNet [9] and Strongest with imperfect CSI for $L = 30$ and 50 communication pairs. Hence, GCRN is more favorable than simple baseline and MLP based methods for power allocation problems in medium or large-scale networks. The performance is also comparable to GNN based WCGCN [3] method, but it requires far fewer training samples than WCGCN.

Next, we test the sum rate performance of GCRN with different value of $d_{\min}$ and $d_{\max}$ for $L = 50$ communication pairs in the area of $1000m \times 1000m$, as shown in Table VII and Fig. 6. We can observe that GCRN achieves good performance compared to WCGCN [15] while using much fewer training

¹The All Active algorithm is a simple baseline algorithm which activates all communication links with $P_{\max}$ power without any knowledge about CSI.

²The Strongest algorithm is another simple baseline algorithm which allocates $P_{\max}$ power to pairs with largest channel gains, while 0 power is set for the rest of the pairs without any knowledge about interfering links.

TABLE VI: AVERAGE SUM RATE PERFORMANCE WITH DIFFERENT NUMBER OF COMMUNICATION PAIRS

Number of Comm. pairs	Proposed GCRN			WCGCN [15] (Perfect CSI)	PCNet [9] (Perfect CSI)	Strongest
	Perfect CSI	$\sigma = 2$	Partial CSI			
10	95.57%	90.57%	89.46%	100%	98.90%	87.10%
30	94.34%	89.63%	88.12%	97.90%	87.40%	82.80%
50	93.82%	88.86%	87.76%	97.50%	79.70%	80.60%

samples and significantly higher performance than the All Active algorithm. It also outperforms the All Active for all the cases presented in the Table VII with imperfect CSI. Fig. 6 compares the scalability performance of the proposed method under different channel conditions such as the perfect CSI, imperfect CSI and partial CSI. Although a slight performance degradation is observed for imperfect CSI, the performance is still satisfactory for different pairwise distances.

TABLE VII: AVERAGE SUM RATE PERFORMANCE WITH DIFFERENT VALUES OF $d_{\min}$ and $d_{\max}$

Pairwise Distance $d_{\min} \sim d_{\max}(m)$	Proposed GCRN			WCGCN [15] (Perfect CSI)	All Active
	Perfect CSI	$\sigma = 2$	Partial CSI		
2m $\sim$ 65m	94.46%	90.17%	88.38%	97.80%	73.65%
10m $\sim$ 50m	93.82%	88.86%	87.76%	97.50%	78.78%
30m $\sim$ 70m	92.15%	88.29%	86.11%	96.50%	60.05%
all 30m	92.57%	88.55%	87.09%	96.80%	69.47%

3) Generalizability to Larger Scales and Higher Densities:

We test the generalizability of the proposed method for wireless networks with larger scales and higher densities. Generalizability testing is different from scalability testing. Different from the scalability testing, where the training is performed every time for each testing scenario, we train our proposed model with tens of communication pairs and test it with unknown network scenes with hundreds of communication pairs, as discussed next.

We first test the generalizability to large-scale networks but with the same user density. We train the GCRN with $L = 50$ communication pairs in $1000m \times 1000m$ wireless network area with $(10m \sim 50m)$ pairwise distance. Then we increase the number of pairs in testing while keeping the user densities fixed (i.e., A^2/L). The results are summarized in Table VIII. The table shows that the performance is promising compared to the WCGCN [15] and outperforms the All Active by a large margin. It can also be observed that the performance is stable as the number of communication pairs increases. The results suggest that the proposed GCRN can generalize to larger problem scales, which is consistent with our permutation equivariance analysis in Proposition 1. Furthermore, the performance is also satisfactory for imperfect CSI and partial CSI, as illustrated in Fig. 7.

Next, we perform the same test but with different user densities. For instance, we first train the proposed model with $L = 50$ communication pairs in $1000m \times 1000m$ network

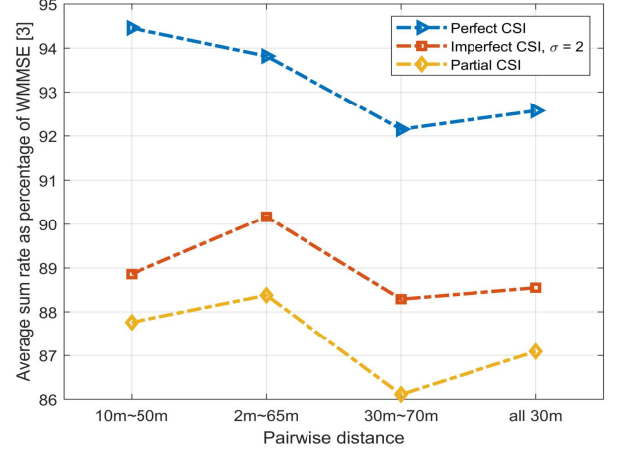


Fig. 6: Scalability of the proposed method under different channel conditions.

TABLE VIII: AVERAGE SUM RATE PERFORMANCE TO LARGER NETWORKS WITH THE SAME USER DENSITY

Number of pairs	Area (m^2)	Proposed GCRN			WCGCN [15] (Perfect CSI)	All Active
		Perfect CSI	$\sigma = 2$	Partial CSI		
200	2000×2000	93.54%	89.76%	87.35%	98.30%	45.92%
400	2828×2828	93.67%	89.82%	88.05%	98.90%	38.47%
600	3464×3464	92.52%	89.64%	87.58%	98.80%	29.62%
800	4000×4000	92.18%	88.26%	86.23%	98.90%	19.53%

area. Then we increase the number of communication pairs in the test dataset while keeping the network area size fixed. The results are summarized in Table IX. As can be seen, the performance is stable till a two-fold increase in the user density, while a good performance is still achieved as compared with WCGCN [15] even when there is a ten-fold increase in the user density with both perfect and imperfect CSI.

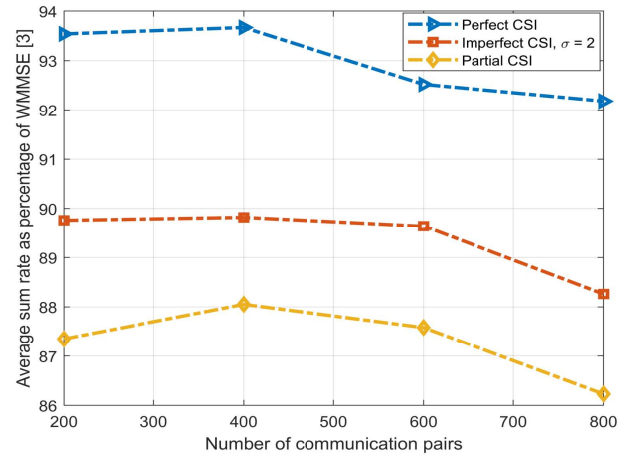


Fig. 7: Generalizability of the proposed method under different channel conditions.

TABLE IX: AVERAGE SUM RATE PERFORMANCE TO LARGER NETWORKS, BUT WITH DIFFERENT USER DENSITY

Number of pairs	Area (m^2)	Proposed GCRN			WCGCN [15] (Perfect CSI)	All Active
		Perfect CSI	$\sigma = 2$	Partial CSI		
100	1000×1000	93.90%	88.91%	87.25%	97.60%	58.38%
200		92.54%	88.21%	87.54%	97.00%	45.43%
400		92.37%	87.62%	86.78%	95.6%	38.25%
500		91.44%	87.06%	86.23%	95.30%	32.46%

C. Unsupervised vs. Supervised Training

As mentioned earlier in Section III, the proposed GCRN model can be trained in a supervised manner using the power allocation of WMMSE [3] scheme as the ground truth. Table X compares the two learning manners (i.e., supervise and unsupervised) on the network area of $1000m \times 1000m$ with $L = 50$ communication pairs for different pairwise distances. As can be seen, supervised learning achieves slightly higher sum rate performance than unsupervised learning for all the cases presented in Table X. However, such an approach is suitable when the labeled training is available, which is computationally challenging especially when the number of nodes is large [9]. In summary, both learning manners have their own pros and cons. Thus, we need to carefully select an appropriate learning approach in practice.

TABLE X: AVERAGE SUM RATE PERFORMANCE FOR UNSUPERVISED AND SUPERVISED LEARNING

Pairwise Distance $d_{\min} \sim d_{\max} (m)$	2m ~ 65m	10m ~ 50m	30m ~ 70m	all 30m
Unsupervised	94.46%	93.82%	92.15%	92.57%
Supervised	95.32%	94.95%	93.56%	93.89%

D. Computational Complexity Analysis

In this subsection, we inspect the computational complexity for the proposed GCRN based method and make a comparison with different ML based state-of-the-arts in the L -user interference channel in Table XI.

The proposed method is implemented in the following sequence: 1) representing the local topology around each communication pair on Riemannian manifolds, and 2) predicting the transmission power by using GCRN. The computational complexity for the communication pair local topology modeling is $\mathcal{O}(L^2)$. On the other hand, for GCRN, we make use of GPU-based implementation of (9) using sparse-dense matrix multiplication [31] and the regression layer is applied row-wise. Then, with a fixed number of iterations, the computational complexity for GCRN based power allocation can be computed as follows:

$$\mathcal{O}(L(|\xi|Ld_1d_2 + |C|)) \approx \mathcal{O}(L^2), \quad (15)$$

where ξ is the number of edges, d_1 and d_2 are the number of feature maps of the two graph convolutional layers and $|C|$ is the total number of trainable parameters of the regression layer.

Thus, the total computational complexity of the proposed GCRN is $\mathcal{O}(L^2)$.

As can be seen in Table XI, the proposed GCRN method has similar computational complexity with MLP [8], DPC [10], and PCNet [9] but the sum rate performance does not deteriorate much for larger networks as observed in these methods. On the other hand, the complexity is higher than WCGCN [15] method, but it requires significantly less training samples than WCGCN to approach the sum rate performance of the benchmark (i.e., WMMSE [3]).

V. CONCLUSION

In this paper, we have introduced a power allocation scheme in large-scale device-to-device (D2D) wireless networks. We aim to reduce the number of training samples required for faster response in dynamic D2D networks. To this aim, we first effectively model the local topology around each communication pair over Riemannian manifolds. Then we have proposed the GCRN method to learn these local topological modelings for predicting wireless power allocation. The proposed method is able to produce stable output with imperfect CSI. It also satisfies the permutation equivalence property, making it possible to generalize to large-scale problems. We have shown that the proposed GCRN based power allocation scheme achieves competitive sum rate performances with only 300 samples as opposed to the need for thousands of training samples by the state-of-the-arts.

REFERENCES

- [1] X. Zhang, H. Zhao, J. Xiong, X. Liu, L. Zhou and J. Wei, "Scalable Power Control/Beamforming in Heterogeneous Wireless Networks with Graph Neural Networks," 2021 IEEE Global Communications Conference (GLOBECOM), 2021, pp. 01-06, doi: 10.1109/GLOBECOM46510.2021.9685457.
- [2] Z. Wang, M. Eisen and A. Ribeiro, "Decentralized Wireless Resource Allocation with Graph Neural Networks," 2020 54th Asilomar Conference on Signals, Systems, and Computers, 2020, pp. 299-303, doi: 10.1109/IEEECONF51394.2020.9443326.
- [3] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2011, pp. 3060–3063.
- [4] C. S. Chen, K.W. Shum, and C.W. Sung, "Round-Robin Power Control for the Weighted Sum Rate Maximisation of Wireless Networks Over Multiple Interfering Links," Eur. Trans. Telecommun., vol. 22, no. 8, pp. 458–470, 2011.
- [5] E. Björnson and E. Jorswieck, "Optimal Resource Allocation in Coordinated Multi-Cell Systems," Foundations and Trends in Communications and Information Theory, vol. 9, no. 2–3, pp. 113–381, 2013.
- [6] Y. Ren, F. Liu, Z. Liu, C. Wang and Y. Ji, "Power Control in D2D-Based Vehicular Communication Networks," in IEEE Transactions on Vehicular Technology, vol. 64, no. 12, pp. 5547–5562, Dec. 2015, doi: 10.1109/TVT.2015.2487365.
- [7] W. Yu, T. Kwon and C. Shin, "Multicell Coordination via Joint Scheduling, Beamforming, and Power Spectrum Adaptation," in IEEE Transactions on Wireless Communications, vol. 12, no. 7, pp. 1–14, July 2013, doi: 10.1109/TWC.2013.052313.121128.
- [8] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, Japan, 2017, pp. 1–6, doi: 10.1109/SPAWC.2017.8227766.
- [9] F. Liang, C. Shen, W. Yu and F. Wu, "Towards Optimal Power Control via Ensembling Deep Neural Networks," in IEEE Transactions on Communications, vol. 68, no. 3, pp. 1760–1776, March 2020, doi: 10.1109/TCOMM.2019.2957482.

TABLE XI: A Comparison of Computational Complexity of Different ML based Power Allocation Methods in the L -user Interference Channel

Methods	MLP [8]	PCNet [9]	DPC [10]	WCGCN [15]	Proposed GCRN
Network Representation Complexity	/	/	Defined by Convolutional Filters	Defined by dimension of Adjacency matrix	$\mathcal{O}(L^2)$
Machine Learning Complexity	/	$\mathcal{O}(L^2)$	$\mathcal{O}(L^2)$	$\mathcal{O}(L)$	$\mathcal{O}(L^2)$
Total Computational Complexity	$= \mathcal{O}(L^2)$	$= \mathcal{O}(L^2)$	$= \mathcal{O}(L^2)$	$= \mathcal{O}(L)$	$= \mathcal{O}(L^2)$

- [10] W. Lee, M. Kim and D. -H. Cho, "Deep Power Control: Transmit Power Control Scheme Based on Convolutional Neural Network," in IEEE Communications Letters, vol. 22, no. 6, pp. 1276-1279, June 2018, doi: 10.1109/LCOMM.2018.2825444.
- [11] N. Naderializadeh, M. Eisen and A. Ribeiro, "Adaptive Wireless Power Allocation with Graph Neural Networks," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 5213-5217, doi: 10.1109/ICASSP43922.2022.9746240.
- [12] B. Li, A. Swami and S. Segarra, "Power Allocation for Wireless Federated Learning Using Graph Neural Networks," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 5243-5247, doi: 10.1109/ICASSP43922.2022.9747764.
- [13] X. Zhang, H. Zhao, J. Xiong, X. Liu, L. Zhou and J. Wei, "Scalable Power Control/Beamforming in Heterogeneous Wireless Networks with Graph Neural Networks," 2021 IEEE Global Communications Conference (GLOBECOM), 2021, pp. 01-06, doi: 10.1109/GLOBECOM46510.2021.9685457.
- [14] N. Naderializadeh, M. Eisen and A. Ribeiro, "Wireless Power Control via Counterfactual Optimization of Graph Neural Networks," 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2020, pp. 1-5, doi: 10.1109/SPAWC48557.2020.9154336.
- [15] Y. Shen, Y. Shi, J. Zhang and K. B. Letaief, "Graph Neural Networks for Scalable Radio Resource Management: Architecture Design and Theoretical Analysis," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 1, pp. 101-115, Jan. 2021, doi: 10.1109/JSAC.2020.3036965.
- [16] A. S. Ibrahim, "Wireless Link Scheduling via Interference-aware Symmetric Positive Definite Connectivity Manifolds," 2021 IEEE International Conference on Communications Workshops (ICC Workshops), 2021, pp. 1-5, doi: 10.1109/ICCWorkshops50388.2021.9473707.
- [17] R. Shelim and A. S. Ibrahim, "Wireless Link Scheduling Over Recurrent Riemannian Manifolds," in IEEE Transactions on Vehicular Technology, vol. 72, no. 4, pp. 4959-4968, April 2023, doi: 10.1109/TVT.2022.3228212.
- [18] R. Shelim and A. S. Ibrahim, "Geometric Machine Learning Over Riemannian Manifolds for Wireless Link Scheduling," in IEEE Access, vol. 10, pp. 22854-22864, 2022, doi: 10.1109/ACCESS.2022.3153324.
- [19] A. S. Ibrahim, "Rethinking Maximum Flow Problem and Beamforming Design Through Brain-inspired Geometric Lens," GLOBECOM 2020 - 2020 IEEE Global Communications Conference, 2020, pp. 1-6, doi: 10.1109/GLOBECOM42002.2020.9321983.
- [20] S. Sra and R. Hosseini, "Conic Geometric Optimization on the Manifold of Positive Definite Matrices," SIAM J. Optimization (SIOPT), vol. 25, no. 1, pp. 713-739, 2015.
- [21] S. Sra, "A New Metric on the Manifold of Kernel Matrices with Application to Matrix Geometric Means," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 144-152. Curran Associates, Inc., 2012.
- [22] I. Nasim and A. S. Ibrahim, "Millimeter Wave Beamforming Codebook Design via Learning Channel Covariance Matrices Over Riemannian Manifolds," in IEEE Access, vol. 10, pp. 119617-119629, 2022, doi: 10.1109/ACCESS.2022.3222032.
- [23] I. Nasim and A. S. Ibrahim, "Relay Placement for Maximum Flow Rate via Learning and Optimization over Riemannian Manifolds," in IEEE Transactions on Machine Learning in Communications and Networking, doi: 10.1109/TMLCN.2023.3309772.
- [24] M.P. do Carmo, "Differential Geometry of Curves and Surfaces: Revised and Updated Second Edition", Dover Books on Mathematics. Dover Publications, 2016.
- [25] J. M. Lee, "Introduction to Riemannian Manifolds," Springer, 2018, doi: 10.1007/978-3-319-91755-9.
- [26] S. Gudmundsson, "An Introduction to Riemannian Geometry," Lund University, 2019.
- [27] M.P. do Carmo, "Riemannian geometry", 2nd ed. Boston, MA: Birkh"auser, 1992.
- [28] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors," Magnetic Resonance in Medicine, vol. 56, no. 2, pp. 411-421, 2006.
- [29] K. You, H.-J. Park, "Re-visiting Riemannian Geometry of Symmetric Positive Definite Matrices for the Analysis of Functional Connectivity", NeuroImage, Volume 225, 2021, 117464, ISSN 1053-8119, <https://doi.org/10.1016/j.neuroimage.2020.117464>.
- [30] Q. Liu, M. Nickel, and D. Kiela, "Hyperbolic Graph Neural Networks", In Advances in Neural Information Processing Systems, pages 8228-8239, 2019.
- [31] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks", In International Conference on Learning Representations (ICLR), 2017.
- [32] D.S. Broomhead and D. Lowe, "Multi-Variable Functional Interpolation and Adaptive Networks", Complex Systems, 2:321-355, 1988.
- [33] T. Poggio and F. Girosi, "Networks for Approximation and Learning", Proceedings of the IEEE, 78(9):1481-1497, 1990.
- [34] M. Belkin and P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering", Advances in Neural Information Processing Systems, 2022, vol. 14, pp. 585-591.
- [35] L. Dodero, H. Q. Minh, M. S. Biagio, V. Murino, and D. Sona, "Kernel based Classification for Brain Connectivity Graphs on the Riemannian Manifold of Positive Definite Matrices," in 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), April 2015, pp. 42-45.
- [36] Z. Dong, S. Jia, C. Zhang, M. Pei, and Y. Wu, "Deep Manifold Learning of Symmetric Positive Definite Matrices with Application to Face Recognition", Proceedings of the AAAI Conference on Artificial Intelligence, 31, 1 (Feb. 2017). DOI:<https://doi.org/10.1609/aaai.v31i1.11232>.
- [37] Z. Gao, Y. Wu, X. Bu, T. Yu, J. Yuan, Y. Jia, "Learning a Robust Representation via a Deep Network on Symmetric Positive Definite Manifolds", Pattern Recognition, Volume 92, 2019, Pages 1-12, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2019.03.007>. (<https://www.sciencedirect.com/science/article/pii/S0031320319301062>)
- [38] N. Keriven, A. Bietti, and S. Vaiter, "Convergence and Stability of Graph Convolutional Networks on Large Random Graphs," In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20). Curran Associates Inc., Red Hook, NY, USA, Article 1806, 21512-21523.
- [39] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", (2017), [Online]. Available: <https://doi.org/10.48550/arXiv.1412.6980>



Rashed Shelim is currently pursuing his Ph.D. degree at the Electrical and Computer Engineering Department at Florida International University (FIU), Miami, FL, USA. He received his BSc degree from North South University, Bangladesh (with summa cum laude) and MSc Degree from RWTH Aachen University, Germany. He is a senior lecturer at the Department of Electrical and Computer Engineering, North South University, Bangladesh (on study leave). Previously, he worked as a summer intern at Nokia Bell Labs, USA, and Core Network Engineer (Packet Switching) at Huawei Technologies Bangladesh LTD. He also worked as a Research Assistant at P3 Communication GmbH, Germany. His principal research interests lie in the field of vehicular communications and rethinking wireless resource allocation by Geometric Machine Learning through the lens of Riemannian geometry.



AHMED S. IBRAHIM is currently an associate professor at the Electrical and Computer Engineering Department at Florida International University (FIU), Miami, FL, USA. He received the B.S. and M.S. degrees in electronics and electrical communications engineering from Cairo University, Cairo, Egypt, in 2002 and 2004, respectively. He received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2009. He is a recipient of the NSF CAREER Award in 2022. Prior to joining FIU, Dr. Ibrahim was an assistant professor at Cairo University, wireless research scientist at Intel Corporation, and senior engineer at Interdigital Communications Inc. Dr. Ibrahim's research interests span various topics of wireless systems including drone-assisted millimeter wave communications, vehicular communications, and rethinking wireless networks through the lens of Riemannian geometry.