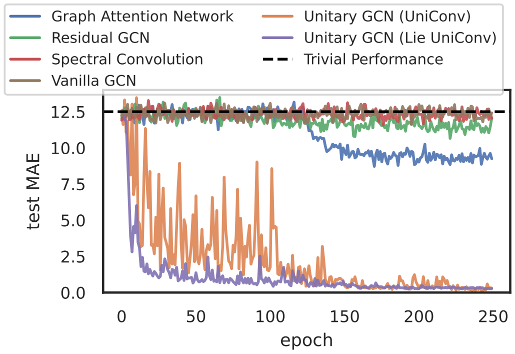

Figure 1: Results for the ring toy model problem with 100 nodes where the unitary GCN uses either UniConv or Lie UniConv layers (both with complex-valued unitary parameterizations). Best performance over networks with 5, 10, and 20 layers is plotted. Other architectures typically perform best with 5 layers and only learn shorter distances.

Weight Parameterization	Method	PEPTIDES-FUNC Test AP ↑	PEPTIDES-STRUCT Test MAE ↓
Unitary (complex-valued)	UniConv	**0.7043 ± 0.0061**	0.2445 ± 0.0009
	Lie UniConv	0.7025 ± 0.0081	0.2461 ± 0.0011
Orthogonal (real-valued)	UniConv	0.7037 ± 0.0053	**0.2433 ± 0.0018**
	Lie UniConv	0.6964 ± 0.0034	0.2471 ± 0.0037

Table 1: Comparison of GCN with UniConv or Lie UniConv layers as well as parameterizations that are real-valued (resulting in orthogonal maps) or complex-valued (resulting in unitary maps) show that UniConv typically performs slightly better. Network have 8 convolution layers with width set to fit within a budget of 500K parameters. Top performer is bolded.

Weight Parameterization	Method	ENZYMES	IMDB	MUTAG	PROTEINS
Unitary (complex-valued)	UniConv	39.13 ± 2.03	**62.80 ± 1.32**	81.28 ± 3.77	**75.97 ± 1.67**
	Lie UniConv	**42.06 ± 3.71**	61.92 ± 1.78	**82.76 ± 4.77**	75.71 ± 2.34
Orthogonal (real-valued)	UniConv	40.56 ± 2.71	62.30 ± 1.45	79.81 ± 3.64	75.06 ± 1.87
	Lie UniConv	41.22 ± 2.86	61.51 ± 1.63	80.47 ± 4.12	74.68 ± 2.13

Table 2: Accuracy of GCN with UniConv or Lie UniConv layers as well as parameterizations that are real-valued (resulting in orthogonal maps) or complex-valued (resulting in unitary maps). Network depths set to 4 layers and the width is set to 256. Results are aggregated over 10 random realizations. Top performer is bolded.

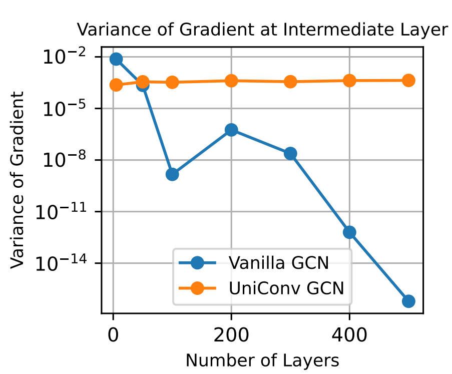

Figure 2: Simple illustration showing vanilla GCNs can suffer from exponentially vanishing or exploding gradients. Here, the variance of the gradients of a convolutional network is shown with the number of convolution layers given in the horizontal axis. The Vanilla GCN suffers from vanishing gradients. Networks are trained over synthetic data on random Erdős–Rényi graphs of 200 nodes. Node features have dimension 32 and set to random i.i.d. Gaussian entries. Variance of the gradients of the intermediate middle layer weights with respect to the MSE loss are shown here.