Costruire un dataset di grafi di spin glass di piccola taglia (# basso di nodi) e studiare come l'apprendimento dell'energia da parte della GNN cambia al variare della connettività del grafo. Ad esempio fissando a 10 il numero di nodi, il numero di edges può essere cambiato da 9 (open spin glass chain) a 45 (fully connected spin glass), magari provando 9, 15, 25, 45. Si costruiscono 4 dataset e si allena la stessa GNN separatamente su questi dataset in modo supervised per fare poi il confronto. (Livello medio)

```
In [ ]: import torch
        from torch_geometric.data import Data
        from torch geometric.utils import to networkx, from networkx
        import networkx as nx
        import matplotlib.pyplot as plt
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
        from matplotlib.offsetbox import AnchoredText
        import torch.nn.functional as F
        from torch geometric.profile import get model size
        from torch.nn import Linear
        from torch_geometric.nn import global_max_pool, GraphConv
        from torch geometric.loader import DataLoader
        import os
In [ ]: # set standard dimensions for the plots
        plt.rcParams['figure.figsize'] = (10, 6)
        plt.rcParams['font.size'] = 14
        plt.rcParams['axes.labelsize'] = 14
        plt.rcParams['axes.titlesize'] = 14
        plt.rcParams['xtick.labelsize'] = 14
        plt.rcParams['ytick.labelsize'] = 14
        plt.rcParams['legend.fontsize'] = 14
In [ ]: def initialize graph(L, periodic=True, n dimensions=2):
            Initialize a graph with L nodes
            G = nx.Graph()
            for i in range(L):
                G.add node(i, spin = np.random.choice([-1, 1]))
            for i in range(L):
                if periodic:
                    for j in range(i+1, i+n dimensions+1):
                        G.add edge(i, j % L, weight = np.random.rand()*2 -1)
                else:
                    for j in range(i+1, i+n dimensions+1):
                        if j < L:
                            G.add edge(i, j, weight = np.random.rand()*2 -1)
            return G
```

1 di 28

```
def plot graph(G, pos=None):
   Plot the graph with nodes colored according to their spin.
   if pos is None:
        pos = nx.circular layout(G)
    colors = ['blue' if G.nodes[i]["spin"] == 1 else 'red' for i in range
    fig = plt.figure()
   nx.draw(G, pos = pos, node color=colors, with labels=True)
   plt.show(fig)
    return fig
def calculate energy(G, h=0 ):
   Calculate the energy of the graph
   E = 0
   E h = 0
   for i, j in G.edges:
        E += G[i][j]["weight"] * G.nodes[i]["spin"] * G.nodes[j]["spin"]
   if h := 0: # exclude the case of h = 0 to avoid floating point error
        E h = h * np.sum([G.nodes[i]["spin"] for i in G.nodes])
    return -E/2 - E h # divide by 2 to avoid double counting
# create a dataset with the energy as label
def create dataset(n nodes=10, h=0, periodic= True, n dimensions=1, n sam
    dataset = []
   graphs = []
    for i in range(n samples):
       G = initialize_graph(n_nodes, periodic=periodic, n_dimensions=n_d
       X = torch.tensor([G.nodes[i]["spin"] for i in G.nodes]).float().v
       graphs.append(G)
       dataset.append(
            Data(x = X,
                edge index = torch.tensor(np.array(G.edges).T, dtype=torc
                edge weight = torch.tensor([G[i][j]["weight"] for i, j in
                y = torch.tensor([calculate energy(G, h)], dtype=torch.fl
    return dataset, graphs
```

2 di 28

```
self.linear1 = Linear(hidden_channels_4, hidden_linear)
                self.linear2 = Linear(hidden linear, num classes)
                self.dropout = torch.nn.Dropout(dropout)
                # we should consider also batch normalization in case of overfitt
                # Optimizer:
                self.optimizer = torch.optim.Adam(self.parameters(), lr=lr, weigh
                # Loss function:
                self.criterion = torch.nn.MSELoss()
            def forward(self, data):
                x, edge_index, edge_weight = data.x, data.edge_index, data.edge_w
                x = F.leaky relu(self.conv1(x, edge index, edge weight))
                x = F.leaky relu(self.conv2(x, edge index, edge weight))
                x = F.leaky relu(self.conv3(x, edge index, edge weight))
                x = F.leaky relu(self.conv4(x, edge index, edge weight))
                x = self.pool(x, data.batch) # [n nodes*n graph of the batch, n f]
                \#x = torch.flatten(x) \# not needed since we are using global max
                #print(x.shape)
                x = self.dropout(x)
                x = F.leaky relu(self.linear1(x))
                out = self.linear2(x)
                return out
            def train step(self, data):
                # Reset gradients
                self.optimizer.zero grad()
                # Passing the node features and the connection info
                prediction = self.forward(data)
                # reshape the prediction to match the shape of the target
                prediction = prediction.view(data.y.shape)
                # Calculating the loss and gradients
                loss = self.criterion(prediction, data.y)
                #print("prediction:", prediction.shape)
                #print("data.y:",data.y.shape)
                loss.backward()
                # Update using the gradients
                self.optimizer.step()
                return loss.item()
In [ ]: def plot_loss( train_losses, test_losses):
            fig = plt.figure()
            epoch = range(1, len(train losses)+1)
```

plt.plot(epoch, train losses, label='Training Loss')

```
plt.plot(epoch, test_losses, label='Test Loss')
    plt.xlabel('Epochs')
    plt.ylabel('Loss')
    plt.title('Losses')
    plt.legend()
    plt.grid(True)
    plt.show()
    return fig
def train model(model, train loader, test loader, num epochs, patience =
    if device is None:
        device = torch.device('cuda' if torch.cuda.is available() else 'c
    # parameters for patience
    best loss = np.inf
    patience counter = 0
    epochs = range(1, num_epochs+1)
    train losses, test losses = [], []
    print('Training ...')
    for epoch in epochs:
        loss per epoch = []
        test loss per epoch = []
        for i, mini batch in enumerate(train loader):
            mini batch.to(device)
            train loss = model.train step(mini batch)
            loss per epoch.append(train loss)
        train losses.append(sum(loss per epoch)/len(loss per epoch))
        with torch.no grad():
            """Evaluate the model on test set"""
            for i, mini batch in enumerate(test loader):
                mini batch.to(device)
                prediction = model.forward(mini batch)
                prediction.to('cpu')
                test_loss = model.criterion(prediction, mini batch.y).ite
                test loss per epoch.append(test loss)
            test losses.append(sum(test loss per epoch)/len(test loss per
            if test losses[-1] < best loss:</pre>
                best loss = test losses[-1]
                torch.save(model.state dict(), model name + '/model.pth')
                patience counter = 0
            else:
                patience counter += 1
                if patience counter == patience:
                    print(f"Early stopping at epoch {epoch}")
                    break
        if verbose:
            print(f"Ep {epoch} \t| Loss {round(train_losses[-1], 5)} \t|
```

```
return train losses, test losses
In [ ]: # set the model name based on the values
        def set model name(n nodes, n_dimensions, h, periodic,n_samples, hidden_c
                            hidden channels 2, hidden channels 3, hidden channels
                            lr, weight decay, dropout, epochs):
            return f"model {n nodes} {n dimensions} {h} {periodic} {n samples} {h
        # translate the model name into values
        def print values from model name(model name, print values=True, return di
            values = model name.split(' ')
            # create a dictionary to store the values
            dict = {"n nodes": values[1], "n dimensions": values[2], "h": values[
                    "hidden channels 1": values[6], "hidden channels 2": values[7
                    "lr": values[11], "weight decay": values[12], "dropout": valu
            if print values:
                for key, value in dict.items():
                    print(f"{key}: {value}")
            if return dict:
                return dict
        def test trained model(model name, test loader, device = None):
            if device is None:
                device = torch.device('cuda' if torch.cuda.is available() else 'c
            values = print values from model name(model name, print values=False,
            trained model = GNN(hidden channels 1= int(values["hidden channels 1"
                        hidden channels 2= int(values["hidden channels 2"]),
                        hidden channels 3= int(values["hidden channels 3"]),
                        hidden channels 4= int(values["hidden channels 4"]),
                        hidden linear= int(values["hidden linear"]),
                        lr= float(values["lr"]), weight decay= float(values["weig")
            trained model.load state dict(torch.load(model name + '/model.pth'))
            trained model.to(device)
            fig = plt.figure()
            with torch.no grad():
                predictions = []
                true values = []
                for i, mini batch in enumerate(test loader):
                    mini batch.to(device)
                    prediction = trained model.forward(mini_batch)
                    predictions.append(prediction.item())
                    true values.append(mini batch.y.item())
                # save the mean squared error
                mse = np.mean((np.array(predictions) - np.array(true values))**2)
                plt.scatter(true values, predictions)
                plt.plot(true_values, true_values, color='red')
                plt.xlabel('True energy')
                plt.ylabel('Predicted energy')
                plt.title('Predicted vs True energy')
```

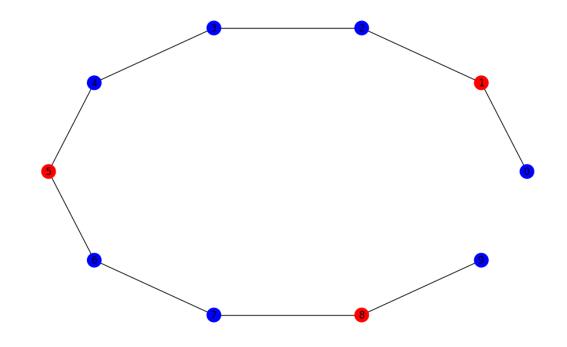
```
at = AnchoredText(f"MSE: {np.round(mse,2)}", frameon=True, loc='u
                 plt.gca().add artist(at)
                 # plot a box similar to the legend to show the MSE
                 plt.grid(True)
                 plt.show()
             return fig, mse
In [ ]: # set the device and a random seed
        torch.manual seed(1022000)
         device = torch.device('cuda' if torch.cuda.is available() else 'cpu')
         print('Device: ', device)
        # set the palette
        palette = sns.color palette("coolwarm", as cmap=True)
       Device: cuda
In [ ]: # set standard parameters for the model
        # parameters for the dataset
        n nodes = 10
        h = 0
        n \text{ samples} = 1000
        # parameters for the model
         num epochs = 30
        batch size = 1
        lr = 1e-4
        weight decay = 1e-4
         dropout = 0.4
         patience = 15
         # parameters for the layers
        hidden channels 1 = 16
         hidden channels 2 = 32
        hidden channels 3 = 32
         hidden channels 4 = 64
        hidden_linear = 32
In [ ]: # create the model
        starting model = GNN(hidden channels 1=hidden channels 1, hidden channels
                     hidden channels 4=hidden channels 4, hidden linear=hidden lin
         # Model size and number of parameters of the model:
         print(44 * '-')
         print("Model:")
        print("Model size (bytes):\t\t ", get_model_size(starting_model))
print("Number of parameters:\t\t ", sum(p.numel() for p in starting_mode
        print(44 * '-')
       Model:
       Model size (bytes):
                                            43862
       Number of parameters:
                                             9457
```

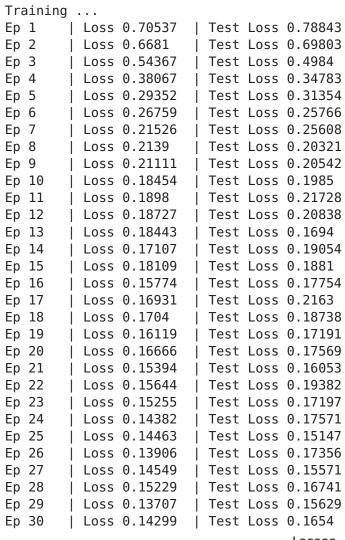
```
In [ ]: # prepare lists to store the results
        list_dimensions = [1,2,3,4,5]
        list_periodic = [False, True]
        list datasets = []
        list graphs = []
        list n connections = []
        list_model_names = []
        list_train_losses = []
        list test losses = []
        list fig losses = []
        list_fig_tests = []
        list mse = []
In [ ]: | for n dimensions in list dimensions:
            for periodic in list periodic:
                # create the dataset
                dataset, graphs = create_dataset(n_nodes=n_nodes, h=h, periodic=p
                list datasets.append(dataset)
                list graphs.append(graphs)
                n connections = len(graphs[0].edges)
                list n connections.append(n connections)
                print(f"n dimensions: {n dimensions}, periodic: {periodic}, numbe
                # plot the first graph
                fig = plot graph(graphs[0])
                # split the dataset into training and test set and create the loa
                train loader = DataLoader(dataset[:int(0.8*n samples)], batch siz
                test loader = DataLoader(dataset[int(0.8*n samples):], batch size
                # copy the model from scratch
                model = GNN(hidden channels 1=hidden channels 1, hidden channels
                    hidden channels 4=hidden channels 4, hidden linear=hidden lin
                model.to(device)
                # set the model name
                model_name = set_model_name(n_nodes, n_dimensions, h, periodic, n
                list model names.append(model name)
                try:
                    os.mkdir(model name)
                except:
                    pass
                # save the graph plot
                fig.savefig(model name + "/graph.png")
                # train the model
                train losses, test losses = train model(model, train loader, test
                list train losses.append(train losses)
                list test losses.append(test losses)
                # save train and test losses
                np.save(model name + '/train losses.npy', train losses)
```

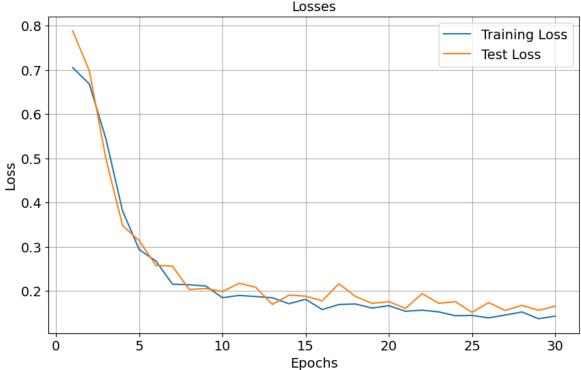
7 di 28

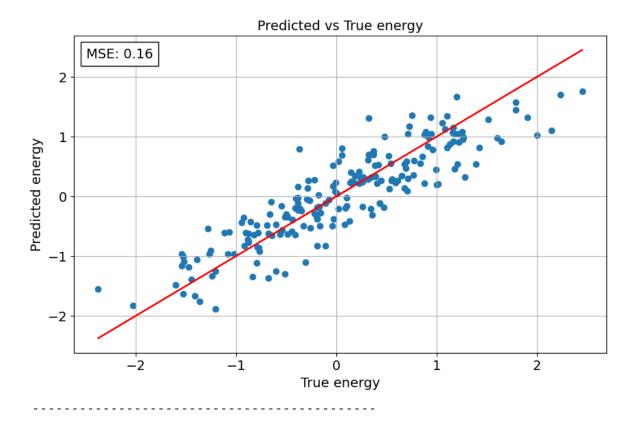
```
np.save(model_name + '/test_losses.npy', test_losses)
# plot the losses
fig losses = plot loss(train losses, test losses)
fig losses.savefig(model name + "/losses.png")
list_fig_losses.append(fig_losses)
# test the trained model
fig, mse = test_trained_model(model_name, test_loader, device=dev
fig.savefig(model name + "/test.png")
list fig tests.append(fig)
# store the mse
list mse.append(mse)
print(44 * '-')
print("\n")
folder name = f"n nodes {n nodes} n dimensions {n dimensions}"
try: os.mkdir(folder name)
except: pass
# move the files to the folder
os.system(f"mv {model name} {folder name}")
```

n_dimensions: 1, periodic: False, number of connections: 9

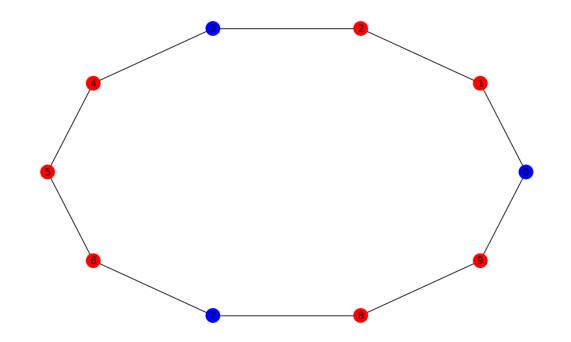


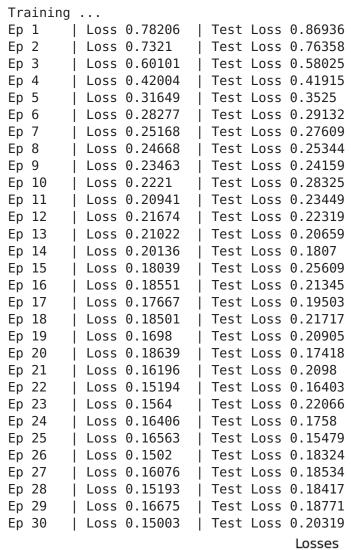


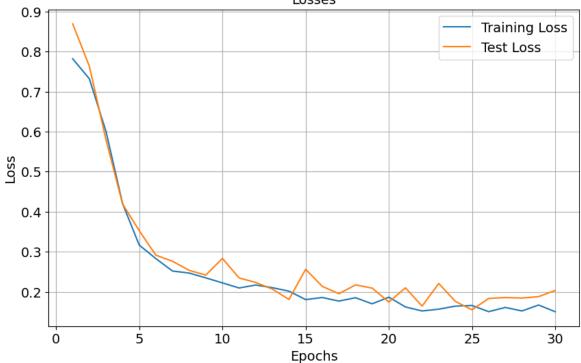


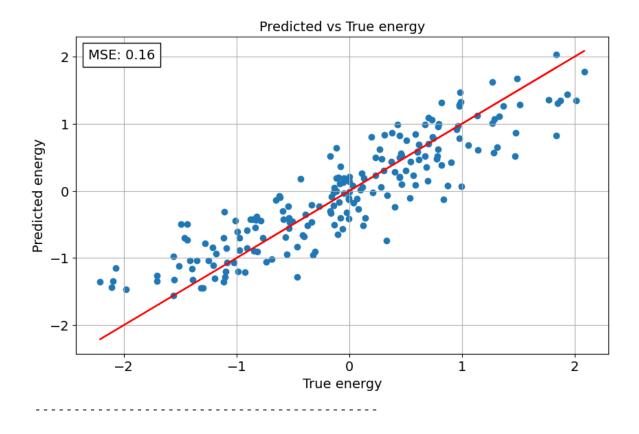


n_dimensions: 1, periodic: True, number of connections: 10

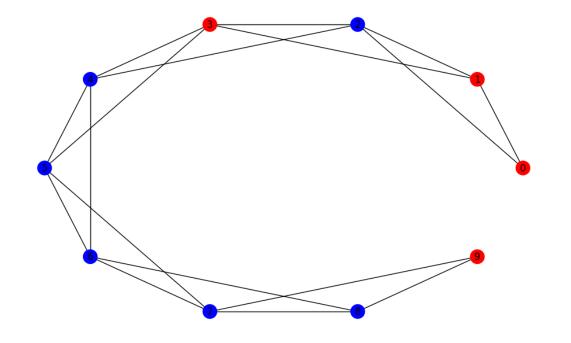


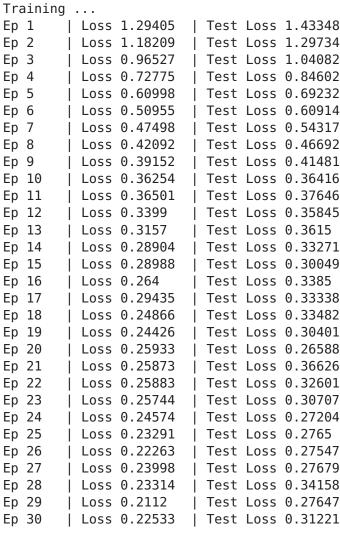


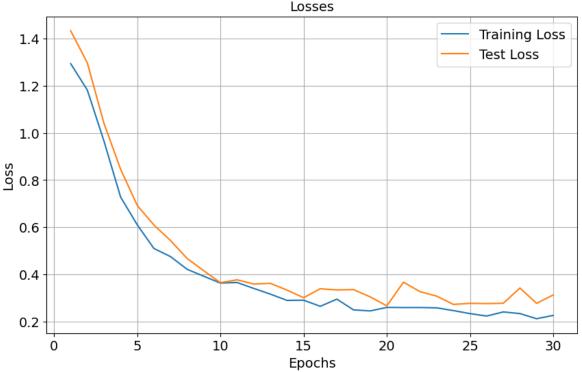


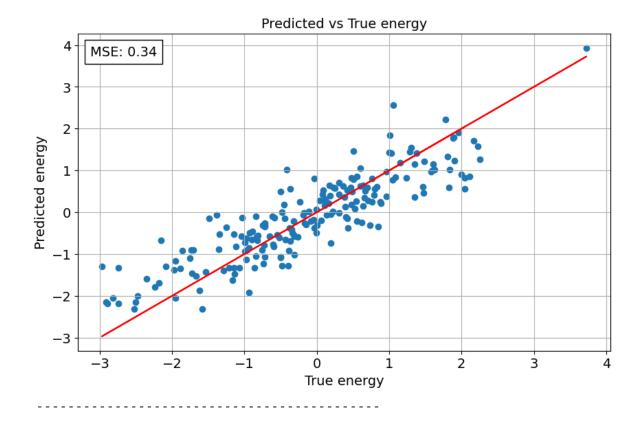


n_dimensions: 2, periodic: False, number of connections: 17

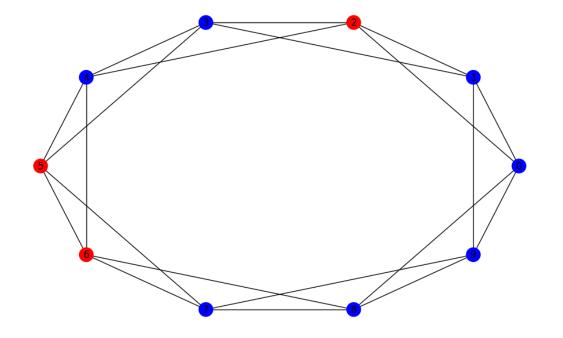


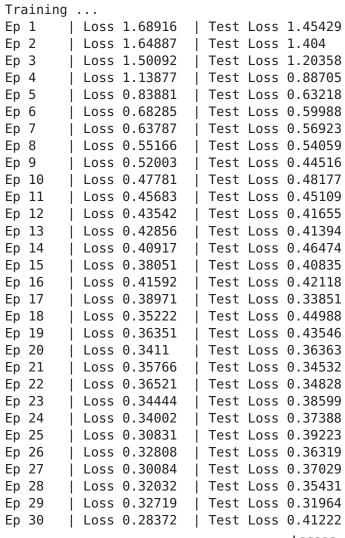


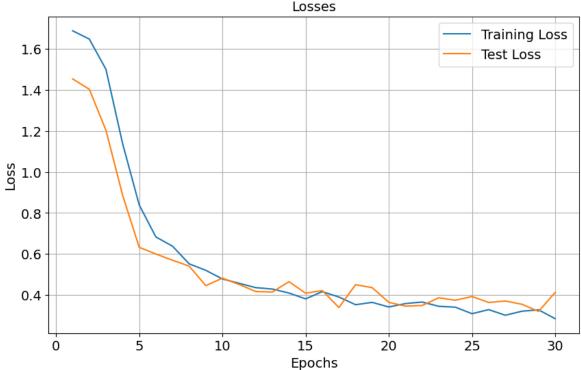


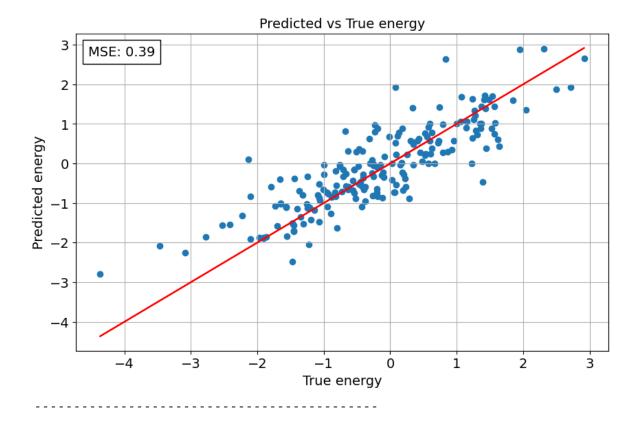


n_dimensions: 2, periodic: True, number of connections: 20

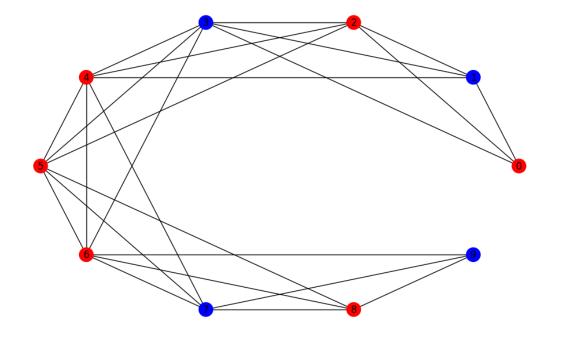


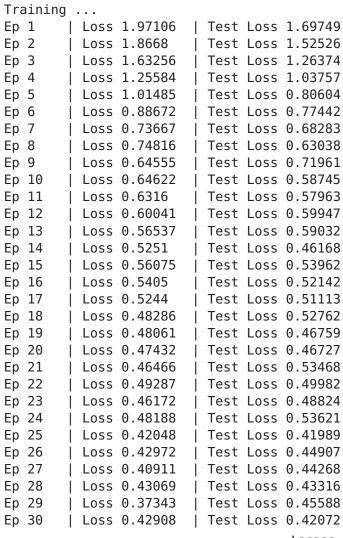


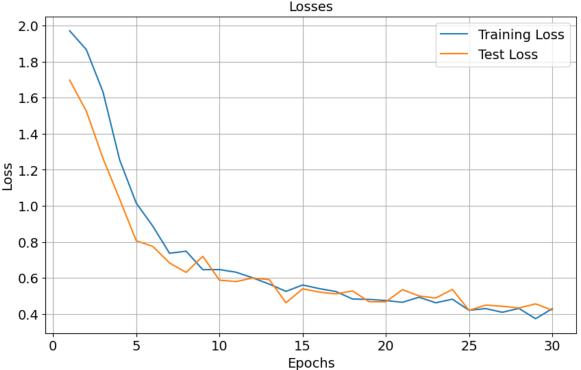


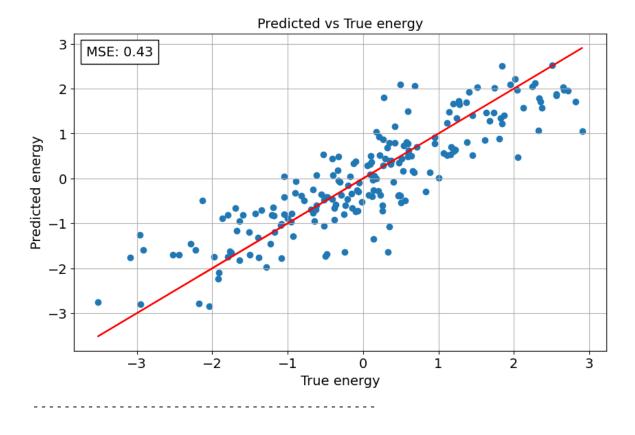


n_dimensions: 3, periodic: False, number of connections: 24

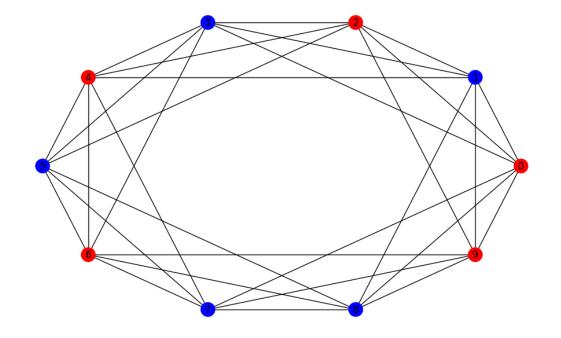


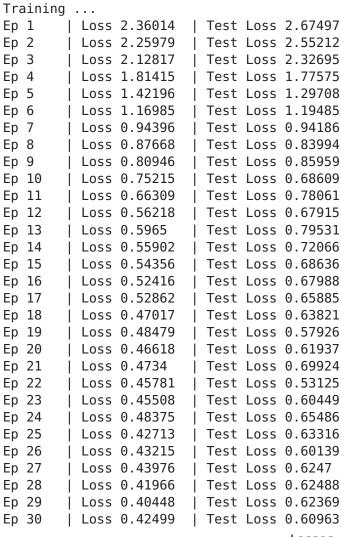


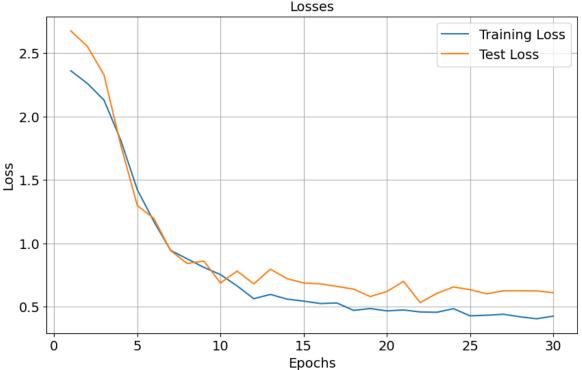


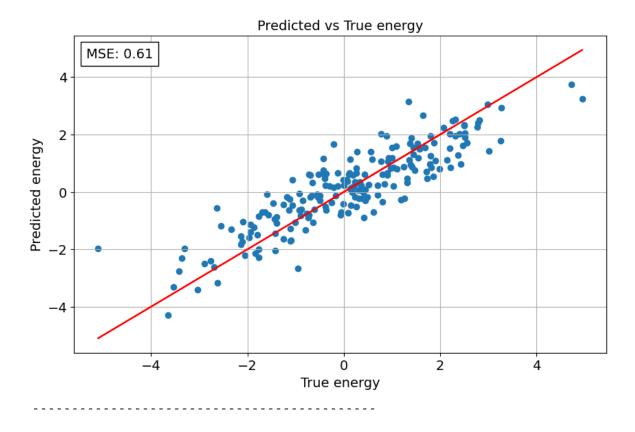


n_dimensions: 3, periodic: True, number of connections: 30

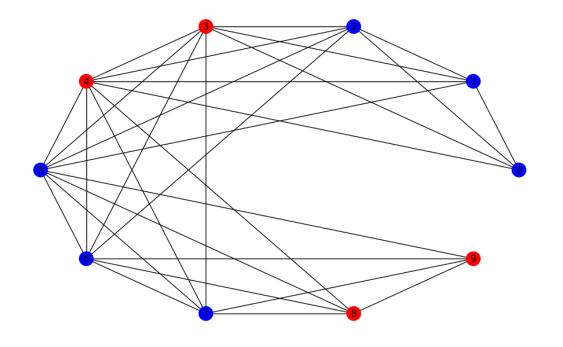


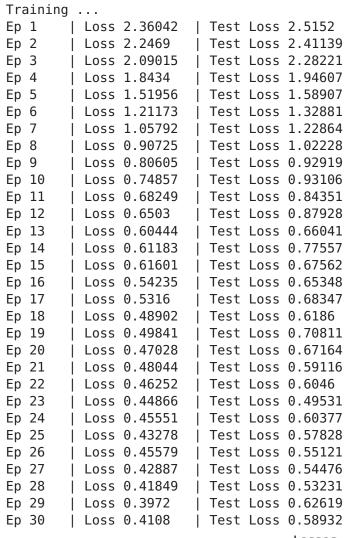


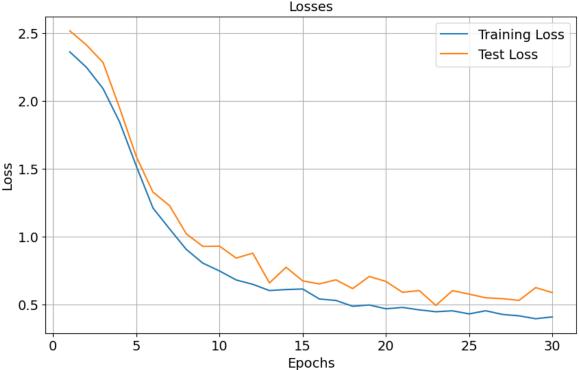


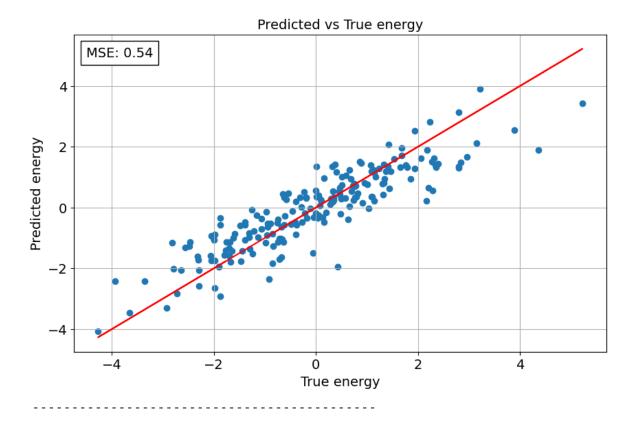


n_dimensions: 4, periodic: False, number of connections: 30

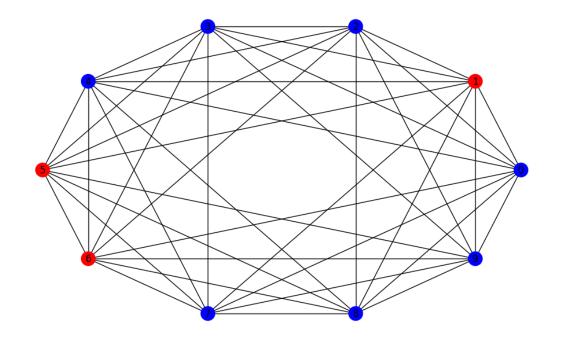


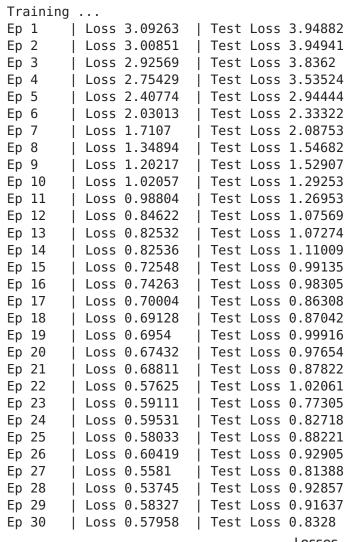


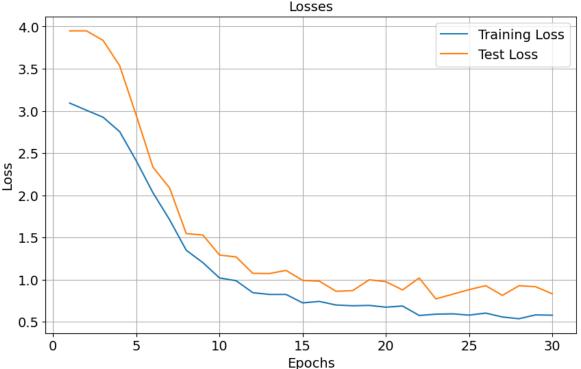


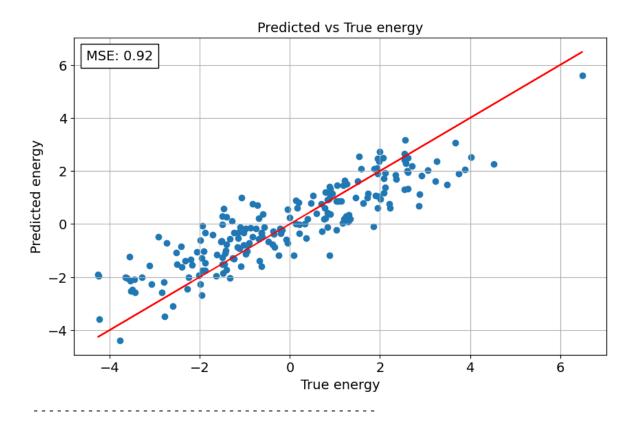


n_dimensions: 4, periodic: True, number of connections: 40

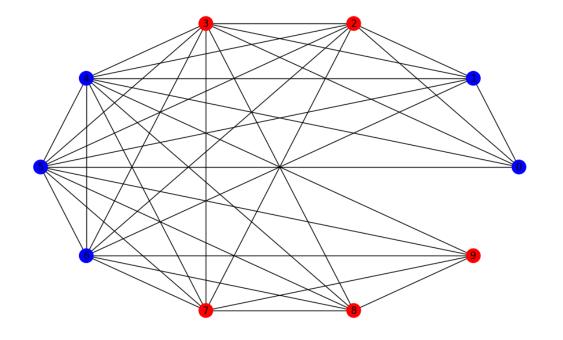


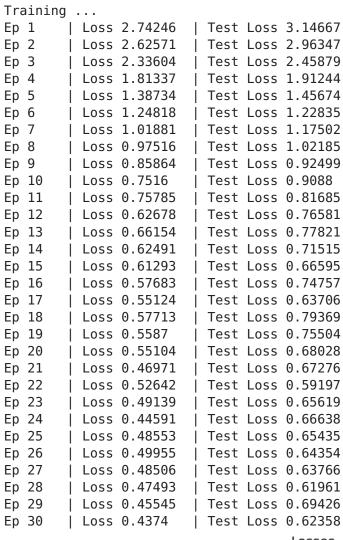


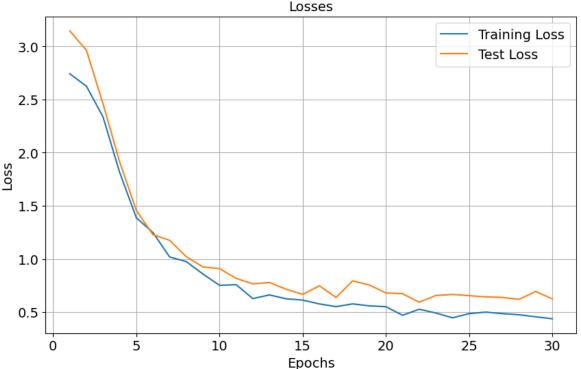


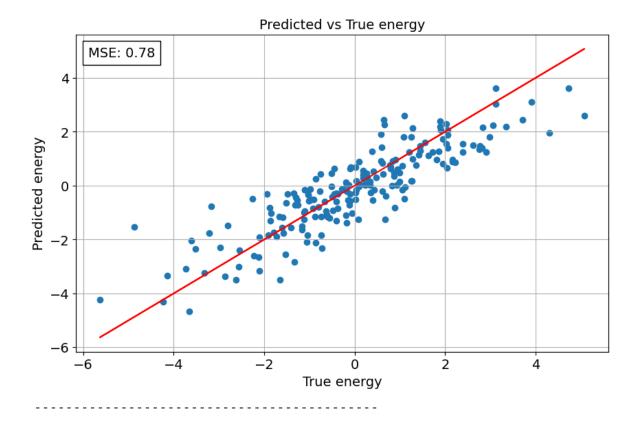


n_dimensions: 5, periodic: False, number of connections: 35









n_dimensions: 5, periodic: True, number of connections: 45

