

图神经网络在准一维伊辛系统中预测磁化的方法

V. Slavin

*B. Verkin Institute for Low Temperature Physics and
Engineering of the National Academy of Sciences of Ukraine,
Nauky Ave., 47, Kharkiv, 61103, Ukraine
email: slavin@ilt.kharkov.ua*

O. Kryvchikov

*B. Verkin Institute for Low Temperature Physics and
Engineering of the National Academy of Sciences of Ukraine,
Nauky Ave., 47, Kharkiv, 61103, Ukraine
email: kryvchikov@ilt.kharkov.ua*

D. Laptev

*B. Verkin Institute for Low Temperature Physics and
Engineering of the National Academy of Sciences of Ukraine,
Nauky Ave., 47, Kharkiv, 61103, Ukraine
email: laptev@ilt.kharkov.ua*

摘要

我们提出了一种基于图的深度学习框架，用于预测准一维伊辛自旋系统的磁性特性。晶格几何被编码为一个图，并通过图神经网络（GNN）以及全连接层进行处理。该模型在蒙特卡罗模拟数据上进行了训练，并准确再现了磁化曲线的关键特征，包括平台、临界转变点和几何挫败的影响。它捕捉到了局部模式和全局对称性，证明了 GNN 可以直接从结构连通性中推断出磁行为。所提出的方法能够高效预测磁化强度，无需额外的蒙特卡罗模拟。

Keywords: 伊辛模型, 蒙特卡罗方法, 图神经网络, 深度神经网络。

I. 介绍

图是一种由顶点集 \mathcal{V} 和边集 \mathcal{E} 定义的数据结构，表示为 $G = (\mathcal{V}, \mathcal{E})$ 。近年来，基于图的方法被广泛应用于分子系统，因为分子可以自然地表示为图。这导致了图神经网络 (GNNs) 的快速发展，GNNs 是一类设计用于处理图形结构数据的深度神经网络 (DNNs) 子类。GNNs 在各种应用中表现出强大的性能，包括预测疏水性¹⁻³ 和毒性⁴⁻⁷ 等理化性质。

在这项工作中，我们将这种方法扩展到统计物理中的问题，特别是到一类一维伊辛模型。系统的哈密顿量由以下给出：

$$\mathcal{H} = J \sum_{\vec{r}_i, \vec{R}} S_{\vec{r}_i} S_{\vec{r}_i + \vec{R}} + h \sum_{\vec{r}_i} S_{\vec{r}_i}, \quad (1)$$

这里， J 表示伊辛自旋 S 之间的耦合，而 h 是外部磁场。我们考虑反铁磁情况下的 $J = 1$ 。相互作用是短程的，满足 $\|\vec{R}\| \ll$ 系统大小。底层晶格几何形状任意但假设为周期性。自旋位置矢量由 $\vec{r}_i = \begin{pmatrix} x \\ y \end{pmatrix}$ 给出，其中横向分量满足 $y \lesssim R$ 。因此该系统是准一维的：它沿着 x 轴延伸，而横向宽度保持有限。

伊辛模型的哈密顿量可以自然地表示为图顶点上的和，其中晶格单元胞的几何结构决定了边的结构。由于这种表述方式，系统可以通过图同构具有等价的表现形式：连接性相同但布局不同的图是无法区分的。这使得哈密顿量能够映射到一维形式，从而便于进一步分析。

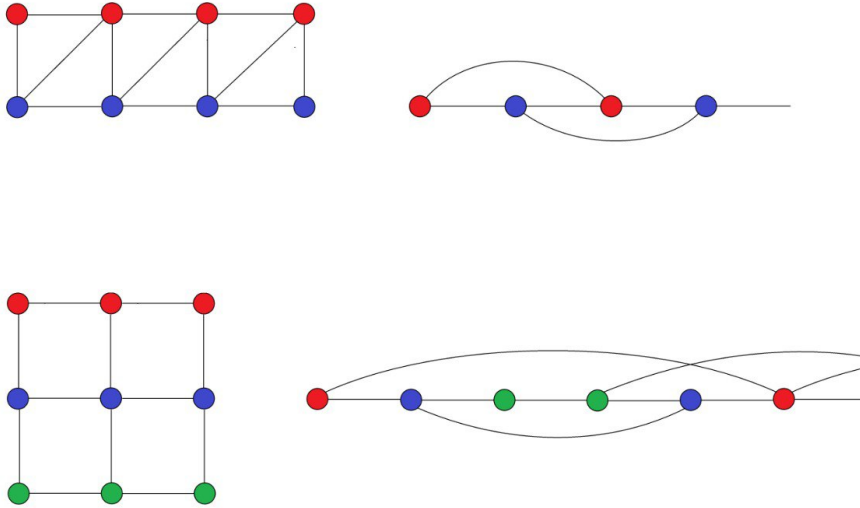


图 1. 两个交互图的同构表示。顶部显示的是一个最近邻交互链，而底部则展示了一个三腿矩形梯子。

映射是通过将哈密顿量划分成 N 个较小的块来实现的，这些块每隔 M 个位置重复一次。总自旋数由 NM 给出。然后，位于位置 \vec{r}_i 的每个自旋被重新标记为 $S_{\vec{r}_i} \rightarrow S_{j+iM}$ ，其中 $j = 0, \dots, M-1$ 是块内索引， $i = 0, \dots, N-1$ 是块索引。

$$\mathcal{H} = \sum_{i=0}^{N-1} \left(\sum_{j \neq k}^M W_{j,k} S_{j+iM} S_{k+iM} + \sum_{j \neq k}^M \tilde{W}_{j,k} S_{j+iM} S_{k+(i+1)M} + \sum_j^M h S_{j+iM} \right). \quad (2)$$

哈密顿量 (2) 完全等价于原始形式 (1), 但映射到一维表示。总能量表示为 N 个相同结构单元的和, 每个单元包含 M 个自旋。在每个单元内, 自旋 S_j 和 S_k 通过耦合矩阵 $W_{j,k}$ 相互作用, 该矩阵编码了块内的相互作用。矩阵 $\tilde{W}_{j,k}$ 描述了位于相邻块中的自旋之间的耦合。

$W_{j,k}$ 和 $\tilde{W}_{j,k}$ 都是对称的, 因为自旋相互作用在交换指标时是不变的。在这项研究中, 我们将 $W_{j,k}$ 和 $\tilde{W}_{j,k}$ 的矩阵元限制为值 0 或 J , 分别对应相应自旋之间是否存在反铁磁相互作用。该形式主义通过允许单位内部和之间的任意耦合模式, 推广了经典的伊辛链模型, 同时仍然不允许跨越远距离块的长程相互作用。

为了获得给定配置的 $W_{j,k}$ 和 $\tilde{W}_{j,k}$ 的磁化曲线, 我们采用经典蒙特卡罗模拟。

哈密顿量 (2) 可以自然地解释为一个无向图, 其中每个顶点对应一个自旋, 边表示成对的相互作用。一旦收集了模拟数据, 我们便使用所得的磁化曲线来训练神经网络。该网络学习从由 $W_{j,k}$ 和 $\tilde{W}_{j,k}$ 定义的交互图到相应磁化行为的映射。经过训练后, 模型可以在不需要额外的蒙特卡罗模拟的情况下预测之前未见过的图形配置的磁化曲线。

II. 模型

为了直接从交互图的结构预测磁化强度——而无需依赖计算成本高昂的蒙特卡罗模拟——我们采用了图神经网络 (GNNs)。现代图神经网络 (GNNs) 非常适合建模如伊辛模型这样的自旋系统, 其中自旋相互作用的基础结构可以自然地表示为一个图。

令 $G = (\mathcal{V}, \mathcal{E})$ 表示输入图, 其中每个边 $e_{ij} \in \mathcal{E}$ 编码自旋 S_i 和 S_j 之间的相互作用 (即非零耦合 W_{ij} 或 \tilde{W}_{ij})。交互图中的每个节点 $u_i \in \mathcal{V}$ 可能携带各种类型的信息, 如结构或拓扑特征。由于实际的自旋状态未知, f_θ 是从图拓扑和编码的边特征映射到宏观可观测量 (如磁化强度) 的过程。换句话说, 我们想要得到如下关系:

$$M(h) = f_\theta(G) \quad (3)$$

其中 $G = (\mathcal{V}, \mathcal{E})$ 是输入图, 而 f_θ 是由可学习权重 θ 参数化的 GNN 模型每个节点特征向量初始化为

$$u_i^{(0)} = \begin{pmatrix} i \\ k_i \end{pmatrix}, \quad (4)$$

其中 i 表示晶胞内的自旋索引 (例如, 1, 2, 3 等), k_i 表示连接到节点 u_i 的相邻自旋数量。边特征 a_{ij} 从相互作用强度 (例如 $W_{ij} = 0$ 或 1) 导出。

GNN 通过一系列消息传递层进行操作。在每一层 k , 节点 u_i 通过一个消息函数从其邻居处聚合信息:

$$m_i^{(k)} = \sum_{i \neq j} \phi_m^{(k)} \left(u_i^{(k-1)}, u_j^{(k-1)}, e_{ij} \right), \quad (5)$$

其中求和遍历 u_i 的邻居, $\phi_m^{(k)}$ 是一个可学习函数, 决定消息如何计算。在 (5) 中的聚合采用对邻居进行均匀求和而不使用注意力机制, 这与图卷积网络 (GCN) 框架⁸一致。

然后使用聚合的消息来更新节点的特征:

$$u_i^{(k)} = \phi_u^{(k)} \left(u_i^{(k-1)}, m_i^{(k)} \right), \quad (6)$$

其中 $\phi_u^{(k)}$ 是一个可学习的更新函数。该过程重复进行 K 层，使自旋状态能够受到交互图中越来越远的邻居的影响。步骤 (5) – (6) 对于每个节点聚合其局部邻域的信息并捕捉每个图的结构特性至关重要。

经过 k 个消息传递层后，特征向量 $u_i^{(k)}$ 编码了其邻域最多 k 跳的信息。GNN 架构中的层数 K 是一个超参数。增加 K 允许网络捕捉图中的长程依赖关系。然而，选择过大值的 K 可能导致过度平滑问题^{9,10}，在这个问题中，图中节点嵌入变得几乎无法区分。

在我们的案例中，节点的数量因图而异，因为每个图代表了一个具有不同自旋数量的磁系统的一种不同的配置。然而，模型必须生成一个固定大小的输出（例如，一个标量磁化值），而不考虑图的大小。为了预测整个系统的属性，在最终层之后应用了一个读出函数 $R(\cdot)$ 以生成一个图级别的表示：

$$u_G = R\left(u_i^{(K)}\right). \quad (7)$$

u_G 是一个固定大小的向量，被称为图嵌入，它编码了来自所有节点 \mathcal{V} 的聚合信息。图嵌入是图的一种连续向量表示形式，捕捉其结构和属性信息。这些嵌入存在于潜在空间中，即一个低维表示空间，在该空间中图相似性对应于几何接近度。使用 (7)，可以定义描述两个图之间相似性的标量参数。这个参数是两个向量之间的角度：

$$S = \frac{(u_{G,1}, u_{G,2})}{\|u_{G,1}\| \|u_{G,2}\|} \quad (8)$$

此度量反映了两个图级别嵌入 u_{G_1}, u_{G_2} 的对齐情况。

若干研究提出了通过嵌入式方法学习和比较图表示的方法。在¹¹中，使用带有注意力机制的神经网络模型来识别重要节点并估计图之间的相似性。¹²中的模型直接使用多尺度卷积进行节点级集匹配，从而绕过了对固定大小图向量的需求。在研究¹³中，应用了一种基于对称注意力的相似度量学习方法来比较图嵌入。

为了获得一个保持置换不变性并捕捉节点嵌入 u_i 分布特性的固定长度向量表示，我们采用了 Set2Set 机制¹⁴。该方法将任意大小的无序集合映射到固定维度的嵌入中，反映了关键统计和结构特征，如聚类、扩散和总体几何形状。结果是，在嵌入空间中，结构相似的图会被投影到相近的位置，即使它们在大小或拓扑上有所不同，而结构不同的图则会产生更多发散的结果。

我们现在提供对该机制如何运作的更详细描述。Set2Set 读出是通过一个由 LSTM 控制的迭代注意力过程在 T 步中计算得出的。长短期记忆 (LSTM) 网络¹⁵⁻¹⁷ 使模型能够捕捉长期依赖关系并从节点特征集合中聚合全局信息，生成一个排列不变向量 u_G 来概括整个图。

每次迭代 t ，都会执行以下操作：LSTM 更新：

$$(q_t, c_t) = \text{LSTM}(r_{t-1}, (q_{t-1}, c_{t-1})), q_0 = 0, \quad c_0 = 0, \quad r_0 = 0 \quad (9)$$

向量 q_t 表示控制器的隐藏状态，并确定当前注意力焦点或图的不同部分的重要性。记忆单元状态 c_t 在迭代过程中保留长期上下文信息。

权重的归一化：

$$\alpha_i^{(t)} = \frac{\exp(u_i^\top q_t)}{\sum_{j=1}^n \exp(u_j^\top q_t)}, \quad \text{for } i = 1, \dots, n \quad (10)$$

在每次迭代步骤中，计算一个注意力权重向量来给图形的不同部分分配重要性分数。

$$r_t = \sum_{i=1}^n \alpha_i^{(t)} u_i \quad (11)$$

节点嵌入然后使用这些权重进行聚合，以生成一个上下文向量，该向量概括了最相关的结构信息。经过 T 次迭代后，获得最终的图嵌入：

$$u_G = [r_1 \| r_2 \| \cdots \| r_T] \quad (12)$$

模型的最终阶段是一个多层感知器 (MLP)，它是深度神经网络 (DNN) 的一个简单子类，由多个全连接层组成。每一层包含神经元，且一层中的所有神经元都与下一层的所有神经元相连。给定一个图级嵌入 u_G ，具有 L 层的 MLP 会将其转换如下：

$$m^{(0)} = f u_G \quad (13)$$

$$m^{(l)} = \sigma \left(W_{\text{MLP}}^{(l)} h^{(l-1)} + b_{\text{MLP}}^{(l)} \right), \quad \text{for } l = 1, \dots, L-1 \quad (14)$$

$$\hat{y} = W_{\text{MLP}}^{(N)} m^{(N-1)} + b_{\text{MLP}}^{(N)} \quad (15)$$

其中， $W_{\text{MLP}}^{(l)}$ 和 $b_{\text{MLP}}^{(l)}$ 分别是第 l 层的权重矩阵和偏置向量，而 $\sigma(\cdot)$ 是一个非线性激活函数。

令 $h = (h_1, h_2, \dots, h_k)$ 是一组离散的外部场值。输出向量 $M(h)$ 表示对应于每个值 h_i 的哈密顿量 (2) 的预测磁化强度。

$$M(h) = \begin{pmatrix} M(h_1) \\ M(h_2) \\ \vdots \\ M(h_k) \end{pmatrix} \quad (16)$$

输出向量的每个组件 \hat{y}_i 对应于离散场 h_i 的预测值，即，

$$M(h_i) = \hat{y}_i, \quad \text{for } i = 1, \dots, k. \quad (17)$$

因此，我们可以从输入图构建以下计算管道到输出：

$$\mathcal{V}, \mathcal{E} \longrightarrow \text{GCN} \longrightarrow \text{Set2Set} \longrightarrow \text{MLP} \longrightarrow \hat{y}_i$$

输出向量 \hat{y}_i 用于定义损失函数 \mathcal{L} ，该损失函数量化了预测磁化曲线与参考磁化曲线之间的差异。通过最小化此损失来训练模型，随着 \mathcal{L} 的减小，性能得到提升。具体而言，采用均方误差 (MSE) 作为损失函数：

$$\mathcal{L} = \frac{1}{k} \sum_{i=1}^k \left(\hat{y}_i - \tilde{M}_i \right)^2 \quad (18)$$

其中 \tilde{M}_i 表示从蒙特卡罗模拟计算出的磁化值，而 \hat{y}_i 是与外部磁场值 h_i 相对应的模型预测。

训练¹⁸是指通过迭代调整模型参数以最小化预定义的损失函数的过程，该损失函数量化了预测输出 \hat{y}_i 与真实目标 \tilde{M}_i 之间的差异。在训练过程中，模型优化其参数 θ 以最小化 \mathcal{L} 。

$$\theta = \{W_{\text{MLP}}, b_{\text{MLP}}, W_{\text{Set2Set}}, b_{\text{Set2Set}}, W_{\text{GCN}}, b_{\text{GCN}}\} \quad (19)$$

在每个训练步骤中，计算损失函数关于模型参数 θ 的梯度：

$$\nabla_{\theta} \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial \theta} \quad (20)$$

在梯度下降中，模型参数通过向与损失梯度相反的方向移动来进行更新。具体来说，如果 θ_t 表示第 t 次迭代的参数且 $\mathcal{L}(\theta_t)$ 是损失，基本更新公式为

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}(\theta_t),$$

其中 η 是学习率。权重矩阵 $W_{\text{MLP}}, W_{\text{Set2Set}}, W_{\text{GCN}}$ 在训练过程中的结构和演变对网络的动力学和泛化能力起着关键作用。最近的研究¹⁹表明，这些矩阵的谱特性，特别是其奇异值的分布，可以强烈影响训练过程。

反向传播^{20,21}是用于计算损失函数相对于模型参数的梯度的算法。它应用微积分的链式法则，有效地将误差信号从输出层向后传递通过整个网络。这使得模型能够确定每个参数对误差的影响，并相应地进行更新。这包括应用链式法则将梯度从输出层向后传播至多层感知机 (MLP)，然后通过 Set2Set 机制，最后到达图卷积网络 (GCN) 层。

$$\text{GCN} \leftarrow \text{Set2Set} \leftarrow \text{MLP}$$

当训练完成后，模型可以预测任何输入图形的磁化曲线 $M(h)$ 。这些预测的质量强烈依赖于训练数据集的代表性和多样性，以及训练过程的效率，包括超参数调整和优化策略。

III. 训练和调优

该数据集是通过收集 80 个不同样本的磁化强度对外部磁场依赖性的方法获得的。每个样本基于包含 1 到 16 个自旋的结构单元，并且单元内的相互作用和单元间的连接模式由特定的输入图定义。将该单元翻译 250 次以形成整个晶格。在每个晶格上，使用 Metropolis 算法执行了蒙特卡罗计算。逆温度为 $\beta = 5.0$ 。磁场间隔使用了 $-4 < h < 4$ 。对于每个磁场的蒙特卡罗计算进行了 12 个循环。每个循环包含 10^6 步骤。最终磁化强度作为循环的平均值进行计算。结果，对于这 80 个样本中的每一个，获得了其磁化强度作为外部磁场和相应输入图的函数。这些数据共同构成了用于训练和评估的数据集。

在我们的问题中，目标输出是磁化曲线 $M(h)$ ，已知它是外部磁场 h 的奇函数且非递减。也就是说，它满足：

$$-M(-h) = M(h), \quad \frac{dM}{d|h|} \geq 0. \quad (21)$$

然而，模型输出函数的对称性性质并未在模型架构中明确编码或强制执行。一种简单的方法是通过增加惩罚项来增强损失函数 \mathcal{L} ，当 $\hat{M}(H)$ 不单调或不小时，这些惩罚项会增加损失：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \lambda_1 \cdot \mathcal{L}_{\text{asym}} + \lambda_2 \cdot \mathcal{L}_{\text{nonmono}}, \quad (22)$$

其中 $\mathcal{L}_{\text{asym}}$ 惩罚 $\hat{M}(H) \neq -\hat{M}(-H)$ ，而 $\mathcal{L}_{\text{nonmono}}$ 惩罚 $\hat{M}(H)$ 减少的区域。然而，经验观察表明，这样的惩罚会降低整体性能并导致回归任务中的准确性损失。

让我们探索一种使用累积增长并带平台控制的替代方法。设 $z_i = \hat{y}_i$ 为所有 $h_i > 0$ 的 MLP 块输出。我们定义一个正向增长率向量 Δ^+ 如下：

$$\Delta_i^+ = \log(1 + e^{z_i}). \quad (23)$$

$\Delta_i^+ > 0$ 对所有 z_i 成立。目标是构造一个依赖于 h 单调的输出向量 $\hat{y}_{\text{right}} \in [0, 1]^k$ ，并且具有可解释的增长步骤。

$$\hat{y}_{\text{right}} = \frac{[f(\Delta_1), f(\Delta_1) + f(\Delta_2), \dots, \sum_{i=1}^k f(\Delta_i)]}{\sum_{i=1}^k f(\Delta_i)}, \quad (24)$$

让我们定义向量 $f(\Delta_i)$ 如下：

$$f(\Delta_i) = \Delta_i^+ \delta_{[\sum \Delta_i^+ > \tau]} \quad (25)$$

克罗内克符号 $\delta_{[\sum \Delta_i^+ > \tau]}$ 引入了一个平台掩码来控制 $y_{\text{right}}(h)$ 的分步行为：

$$\delta_{[z_i > \tau]} = \begin{cases} 1, & \text{if } \sum \Delta_i^+ > \tau, \\ 0, & \text{otherwise.} \end{cases} \quad (26)$$

其中 τ 是高原阈值的超参数，求和 $\sum \Delta_i^+$ 在当前高原上进行。

为了确保对称性，我们定义左侧减少为：

$$\hat{y}_{\text{left}} = -\text{flip}(y_{\text{right}}), \quad (27)$$

最终的平滑函数将会是这样的：

$$\hat{y}_{\text{smooth}} = [y_{\text{left}}, 0, y_{\text{right}}]. \quad (28)$$

这确保了所得函数是奇函数且单调。

最终的预测输出是一个加权和：

$$\hat{y} = (1 - \alpha) \cdot \hat{y}_i + \alpha \cdot \hat{y}_{\text{smooth}}, \quad (29)$$

允许模型结合学习到的尖锐响应与平台感知的累积行为。即使一个小的 $\alpha \approx 0.05$ 也能产生更平滑的结果和更具相关性的 $M(h)$ 曲线。

在机器学习中，超参数如 α 是外部配置值，它们不是从训练数据中学到的，而是在训练过程开始之前设置的。它们定义了模型的结构并影响训练动力学。选择合适的超参数对于实现良好的性能和模型泛化至关重要。

对于 GCN→Set2Set→MLP 架构²²，关键超参数包括：学习率、MLP 中的层数、每层 MLP 的隐藏单元数、GCN 层的数量、激活函数的选择、批处理大小和训练周期数、平台阈值、混合系数。找到超参数的最佳组合是一项非平凡的任务。一种常用的方法是网格搜索²³，其中对每个超参数都穷尽探索预定义的一组可能值。模型针对每种组合进行训练，选择验证性能最佳的一种。然而，随着超参数的增加，网格搜索变得效率低下，尤其是对于 GCN+Set2Set+MLP 这样的复合模型。在复合模型的情况下，手动调整超参数可能更有效，因为它允许根据性能立即反馈进行顺序调整。此外，虽然损失函数衡量了预测输出与目标输出之间的差异，但它通常不能反映预测的物理合理性。

IV. 结果

一个单独的测试数据集是按照与训练数据集相同的程序构建的。每个条目由一个输入交互图和相应磁化曲线 $M(h)$ 组成。重要的是，测试集中包含的图形完全不同于训练期间使用的那些图形，确保模型没有接触过它们的结构特征，并允许对泛化性能进行无偏评估。

为了定量评估通过蒙特卡罗模拟获得的预测磁化曲线与参考磁化曲线之间的吻合程度，采用了以下复合误差指标：

$$E = \frac{1}{2} \cdot \frac{1}{k} \sum_{i=1}^k \left| M_i^{\text{MC}} - M_i^{\text{pred}} \right| + \frac{1}{2} \cdot \frac{1}{k} \sum_{i=1}^k \left| \frac{dM^{\text{MC}}}{dh_i} - \frac{dM^{\text{pred}}}{dh_i} \right| \quad (30)$$

该比较是在 k 个离散外部磁场值上进行的，这些值如 (17) 中所定义。此指标同时量化了磁化值的绝对偏差以及局部斜率的差异，从而捕捉到了预测曲线中的垂直和结构误差。

在对之前未见过的测试图进行评估时，该模型显示了定性准确，并且在许多情况下与蒙特卡罗结果定量精确一致。观察到的最小误差为 $E = 0.045$ 。试图通过聚类及相关分析技术将预测准确性与图形的结构属性相关联并未产生任何一致模式。因此，我们无法可靠地预测哪些图配置在我们的模型上表现更好。最大的误差 $E = 0.389$ 观察到了两腿矩形格子中。这种相对较差的表现可能归因于该几何形状的结构规则性和低复杂性，这可能会限制 GNN 提取信息交互特征的能力。或者，它可能是由训练数据集中的不平衡或稀疏引起的。

为了调查数据集的结构组织，我们采用了 t 分布随机邻域嵌入 (t-SNE) 算法，这是一种非线性降维技术，可以将高维数据投影到低维空间中，同时保留局部相似关系²⁴。应用于图表示，t-SNE 允许对聚类倾向和图实例之间的结构相似性进行定性评估。

如图 2 所示，二维 t-SNE 嵌入揭示了四个松散分离的区域，每个区域大致对应于特定类型的图形结构。浅蓝色区域主要包含类似矩形的晶格，这些晶格表现出高度的规则性和对称性。粉色区域主要对应于链状且非挫败的晶格。红色和深蓝色区域主要由具有更复杂连接性的挫败晶格组成，包括奇数环和侧耦合等模式。

然而，应该注意到在这些区域中的某些程度上可能仍然存在挫败感，并且 t-SNE 聚类数据的确切标准并未完全确定。观察到的聚类反映了高维输入空间中的模式，但并不一定对应于唯一的物理分类。尽管如此，这种可视化表明学习到的图嵌入捕捉到了可能与预测磁性行为相关的结构差异。

尽管存在这些限制，该模型仍能一致地捕捉到大多数测试案例中磁化曲线的基本定性特征，包括磁化平台的数量和高度，以及过渡点的位置。

结构信息系统地反映在模型生成的预测中。具有更高对称度的图通常会产生更平滑的磁化分布，而那些连接稀疏或不规则的图则表现出更突然的过渡。

这些趋势与一维自旋系统已建立的理论结果一致。先前的研究表明，非挫折构型倾向于表现出明确的磁化台阶和尖锐的磁场诱导转变²⁵⁻²⁷。相比之下，挫折几何结构——特别是那些涉及奇数长度循环上的反铁磁耦合或显示出组分无序的结构——导致更复杂的磁化行为，包括平台平坦化、延迟饱和以及总磁化的抑制²⁸⁻³²。值得注意的是，所提出的模型仅基于结构输入就成功捕捉到了这些定性差异，突显了其提取并内化控制低维磁现象的基本物理特征的能力。

此外，该网络对已知会引入挫败感的局部结构基序表现出敏感性，例如侧面耦合和奇数校验周期，这反过来又调制了磁化曲线的全局形态。

一旦训练完成，该模型便能够快速预测由方程 (2) 描述的任意一维自旋链拓扑的磁化行为，从而避免了重复的蒙特卡罗采样。这使得计算开销显著减少，同时保持物理上一致的输出。

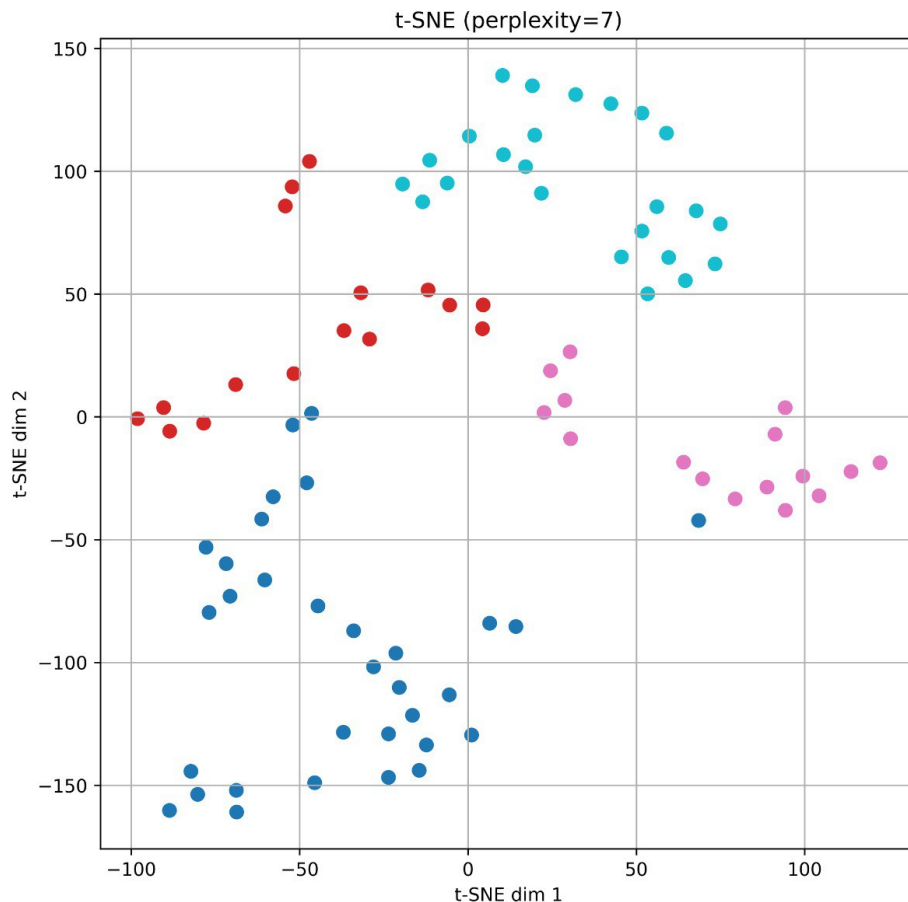


图 2. t-SNE 投影展示了数据集中的四个不同聚类。浅蓝色区域对应矩形晶格，粉色区域对应链状非挫败晶格，红色和蓝色区域代表不同类型的成绩晶格。

本研究采用了反铁磁伊辛模型与 $J = 1$ 作为最小但非平凡的设置，以评估基于图的神经网络在编码和预测磁响应函数方面的有效性。未来改进的几个途径自然随之而来。加入表示耦合常数大小和符号的边级属性可以提高模型的表现力。集成注意力机制可能进一步使网络能够有选择地强调物理上显著的子结构。将框架扩展到更高维度的自旋系统将允许研究更复杂的几何挫折和相行为形式。此外，推广到包括 Potts、Heisenberg 和量子变体在内的替代自旋模型将进一步扩大该方法的应用范围，并通过基于结构的学习促进对磁性的深入探索。

¹ Shang, C., Liu, Q., Chen, K., Sun, J., Lu, J., Yi, J. et al. Edge attention-based multi-relational graph convolutional networks. *ArXiv:1802.04944* (2018).

² Wang, X., Li, Z., Jiang, M., Wang, S., Zhang, S. & Wei, Z. Molecule property prediction based on spatial graph embedding. *Journal of Chemical Information And Modeling*. **59**, 3817-3828 (2019).

³ Chen, B., Bécigneul, G., Ganea, O., Barzilay, R. & Jaakkola, T. Optimal transport graph neural networks. *ArXiv:2006.04804* (2020).

⁴ Xu, Y., Pei, J. & Lai, L. Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction. *Journal of Chemical Information And Modeling*. **57**, 2672-2685

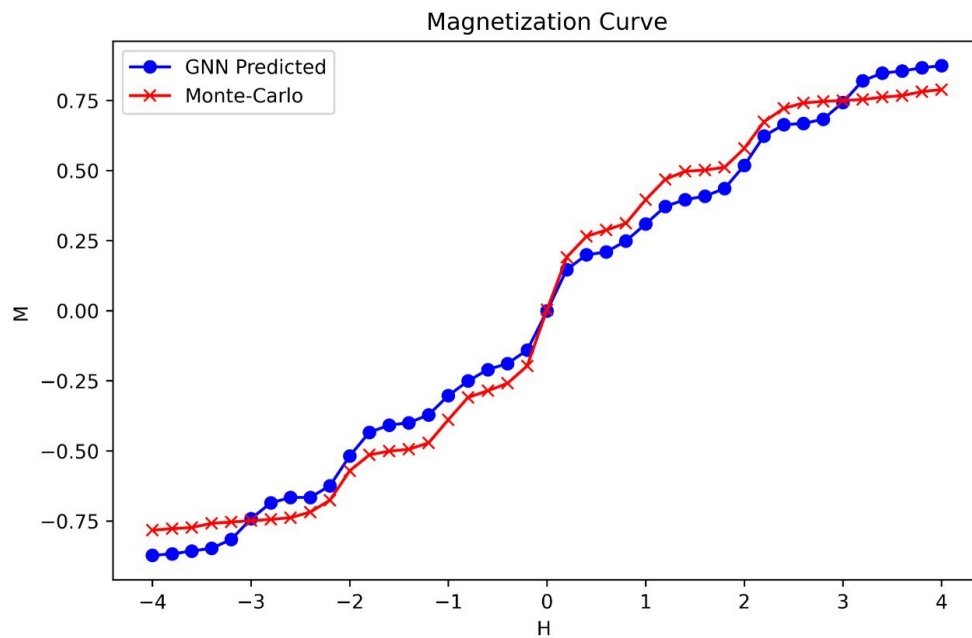


图 3. 磁化强度对外部磁场的依赖性。测试的蒙特卡罗（红色）和预测值（蓝色）。 $E=0.17$ 。

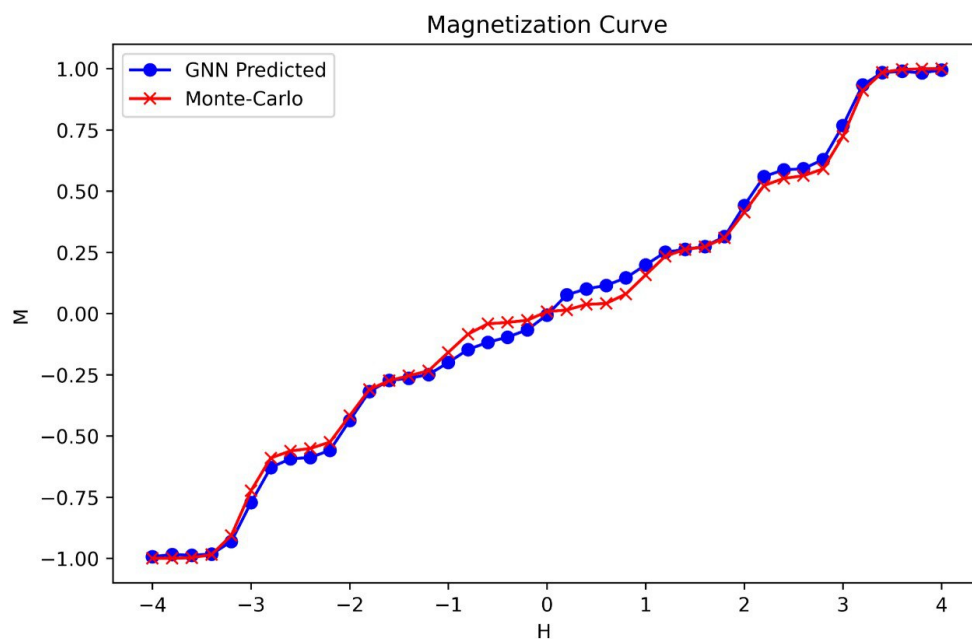


图 4. 磁化率对外部磁场的依赖性。测试的蒙特卡罗（红色）和预测值（蓝色）。 $E=0.08$ 。

(2017).

- ⁵ Withnall, M., Lindelöf, E., Engkvist, O. & Chen, H. Building attention and edge message passing neural networks for bioactivity and physical-chemical property prediction. *Journal of Cheminformatics*. **12** (2020).
- ⁶ Yuan, H. & Ji, S. StructPool: Structured Graph Pooling via Conditional Random Fields. *International Conference on Learning Representations* (2020).
- ⁷ Hu, W., Liu, B., Gomes, J., Zitnik, M., Liang, P., Pande, V. et al. Strategies for Pre-training graph neural

- networks. *ArXiv Preprint ArXiv:1905.12265* (2019).
- ⁸ Kipf, T. & Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *ArXiv:1609.02907* (2016).
 - ⁹ Li, Q., Han, Z. & Wu, X. Deeper insights into graph convolutional networks for semi-supervised learning. *Proceedings of The AAAI Conference on Artificial Intelligence*. **32** (2018).
 - ¹⁰ Oono, K. & Suzuki, T. Graph neural networks exponentially lose expressive power for node classification. *ArXiv:1905.10947* (2020).
 - ¹¹ Bai, Y., Ding, H., Bian, S., Chen, T., Sun, Y. & Wang, W. SimGNN: A neural network approach to fast graph similarity computation. *Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM)*. 384–392 (2019). DOI: 10.1145/3289600.3291015.
 - ¹² Bai, Y., Ding, H., Gu, K., Sun, Y. & Wang, W. Learning-based efficient graph similarity computation via multi-scale convolutional set matching. *Proceedings of the AAAI Conference on Artificial Intelligence*. **34**(4), 3219–3226 (2020). DOI: 10.1609/aaai.v34i04.5720.
 - ¹³ Tao, T., Wang, Q., Ruan, Y., Li, X. & Wang, X. Graph embedding with similarity metric learning. *Symmetry*. **15**(8), 1618 (2023). DOI: 10.3390/sym15081618.
 - ¹⁴ Vinyals, O., Bengio, S. & Kudlur, M. Order matters: Sequence to sequence for sets. *ArXiv:1511.06391* (2016).
 - ¹⁵ Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Computation*. **9**, 1735-1780 (1997).
 - ¹⁶ Gers, F. A., Schmidhuber, J. & Cummins, F. Learning to forget: continual prediction with LSTM. *Proc. Ninth International Conference on Artificial Neural Networks (ICANN '99)*. vol.2, 850 – 855 (IEE, Edinburgh, UK, 1999). DOI: 10.1049/cp:19991218.
 - ¹⁷ Greff, K., Srivastava, R., Koutnik, J., Steunebrink, B. & Schmidhuber, J. LSTM: A search space odyssey. *IEEE Transactions on Neural Networks And Learning Systems*. **28**, 2222-2232 (2016).
 - ¹⁸ Nielsen, M.A. *Neural Networks and Deep Learning*. Determination Press, 2015.
 - ¹⁹ Pastur, L. & Slavin, V. On random matrices arising in deep neural networks: General I.I.D. case. *Random Matrices: Theory and Applications* **12**(01), 2250046 (2023). Available at: <https://doi.org/10.1142/S2010326322500460>
 - ²⁰ Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. *Nature*. **323**, 533-536 (1986).
 - ²¹ Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. MIT Press, 2016. Available at: <http://www.deeplearningbook.org>
 - ²² Wang, Y., Li, Z. & Farimani, A. Graph Neural Networks for Molecules. *arXiv:2209.05582* (2022).
 - ²³ Zahedi, L., Mohammadi, F., Rezapour, S., Ohland, M. & Amini, M. Search Algorithms for Automated Hyper-Parameter Tuning. *ArXiv:2104.14677* (2021).
 - ²⁴ Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research*. **9**, 2579-2605 (2008).
 - ²⁵ Zvyagin, A., Slavin, V., Zvyagina, G. & Piryatinskaya, V. Intersite spin nematic ordering in spin chain materials: Behavior of observables. *Physical Review B*. **111**, 104415 (2025).
 - ²⁶ Zvyagin, A. & Slavin, V. Intersite spin nematic ordering in the spin-chain system. *Physical Review B*. **109**, 104409 (2024).
 - ²⁷ Zvyagin, A., Slavin, V. & Zvyagina, G. Manifestation of spin nematic ordering in the spin-1 chain system. *Physical Review B*. **107**, 134421 (2023).
 - ²⁸ Laptiev, D., Kryvchikov, O., Savin, Y. & Slavin, V. Magnetic properties of the frustrated Ising chain. *Low Temperature Physics*. **50**, 158-166 (2024).
 - ²⁹ Slavin, V. & Cheranovskii, V. Magnetic properties of 1D spin systems with compositional disorder of three-spin structural units. *Zeitschrift Für Naturforschung B*. **76**, 797-802 (2021).
 - ³⁰ Kryvchikov, O. Effective nonfrustrated model of the antiferromagnetic spin-1/2 ladder. *Low Temperature Physics*. **48**, 200-205 (2022).

- ³¹ Cheranovskii, V., Slavin, V., Ezerskaya, E., & Dronkowski, R., Magnetic Properties of Quasi-One-Dimensional Crystals Formed by Graphene Nanoclusters and Embedded Atoms of the Transition Metals. *Crystals*. **9**(5), 251 (2019).
- ³² Slavin, V., Thermodynamics of a one-dimensional generalized Wigner crystal. *Low Temperature Physics*. **29**, 674 – 678 (2003).