Predicting and Interpreting Energy Barriers of Metallic Glasses with Graph Neural Networks

Haoyu Li [*1] **Shichang Zhang** [*1] **Longwen Tang** [1] **Mathieu Bauchy** [1] **Yizhou Sun** [1]

Abstract

Metallic Glasses (MGs) are widely used materials that are stronger than steel while being shapeable as plastic. While understanding the structure-property relationship of MGs remains a challenge in materials science, studying their energy barriers (EBs) as an intermediary step shows promise. In this work, we utilize Graph Neural Networks (GNNs) to model MGs and study EBs. We contribute a new dataset for EB prediction and a novel Symmetrized GNN (SymGNN) model that is E(3)-invariant in expectation. SymGNN handles invariance by aggregating over orthogonal transformations of the graph structure. When applied to EB prediction, SymGNN are more accurate than molecular dynamics (MD) local-sampling methods and other machine-learning models. Compared to precise MD simulations, SymGNN reduces the inference time on new MGs from roughly **41 days** to **less than one second**. We apply explanation algorithms to reveal the relationship between structures and EBs. The structures that we identify through explanations match the medium-range order (MRO) hypothesis and possess unique topological properties. Our work enables effective prediction and interpretation of MG EBs, bolstering material science research. [1]

1. Introduction

Metallic glasses (MGs) combine good properties of metals and plastics in one material, making them stronger than steel while being shapeable as plastic (Schroers et al., 2011). Their extensive applications span various industries including aerospace, sports equipment, luxury goods, biomedical devices, and many more (Trexler & Thadhani, 2010). The unique properties of MGs lie in their non-crystalline amorphous atomic structure, which sets them apart from the crystalline structure found in traditional metals (Trexler & Thadhani, 2010; Bansal & Doremus, 2013). Despite extensive research on MGs, the details of their structure-property relationship are still not well understood (Starr et al., 2002; Ding et al., 2014; Patinet et al., 2016; Cao et al., 2018).

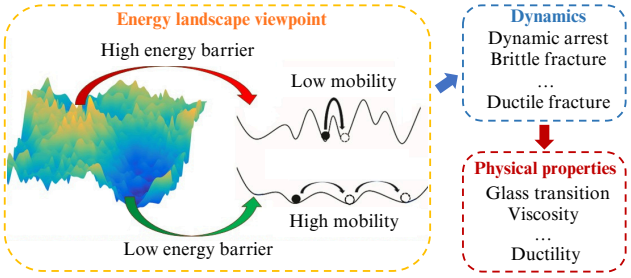

Figure 1. EBs represent mobility, which can further influence the MG dynamics and their physical properties.

One promising approach for studying the structure-property relationship of MGs is through a special property called *Energy Barrier (EB)*. EBs describe the local roughness of the energy landscape by comparing the average energy difference around an atom's local neighbors. Many studies have shown that understanding EBs can act as an important intermediary step for studying the MG physical properties (Debenedetti & Stillinger, 2001; Yu et al., 2012; Tang et al., 2021). As shown in Figure 1, EBs represent mobility, which can influence the MG dynamics and further their physical properties like glass transition and ductility (Berthier & Biroli, 2011; Kirchner et al., 2022). However, the precise simulation of EBs is challenging and often requires time-consuming computation (Barkema & Mousseau, 1996; Mousseau et al., 2012; Jay et al., 2022). For example, even with a high-performance computing (HPC) cluster and the advanced Activation-Relaxation Technique nouveau (ARTn) (Cances et al., 2009), calculating EBs for an MG system with 3,000 atoms can take 41 days.

Given the usefulness and computational difficulty of EBs, we explore machine learning (ML) approaches to efficiently

[*]Equal contribution [1]University of California, Los Angeles, CA, USA. Correspondence to: Haoyu Li <haoyuli02@ucla.edu>, Shichang Zhang <shichang@cs.ucla.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

[1]Code for this project is available at https://github.com/haoyuli02/SymGNN

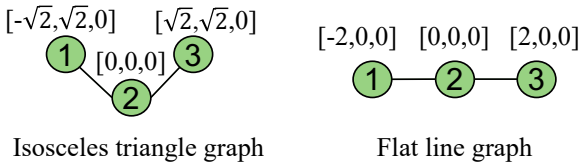

Figure 2. Example graphs demonstrating model expressiveness. SchNet cannot distinguish the embeddings of node 1 in these two graphs but SymGNN can.

predict them from MG atomic structures. Similar to recent ML investigations on glassy systems (Bapst et al., 2020; Reiser et al., 2022), we phrase the EB prediction problem as a graph ML problem and solve it using Graph Neural Networks (GNNs). Under this formalization, atoms become nodes in a graph, and edges are constructed between nearby nodes to represent the atomic structure. Atom types are used as node features. Displacement vectors constructed from 3D node coordinates are used as edge features. Then EB prediction becomes a node regression task on graphs.

We simulate MG systems and employ ARTn to calculate some EBs as training labels. Given the challenge of collecting data, a more data-efficient model with a stronger inductive bias is desired. In particular, the EB prediction problem exhibits E(3)-invariance, i.e., invariance to graph structure transformations including translations, rotations, reflections, and their combinations. We aim for a GNN that can handle such invariance, but general message-passing-based GNNs like GCN (Kipf & Welling, 2017) cannot. Some specially designed models are E(3)-invariant (Schütt et al., 2017; Lu et al., 2019; Gasteiger et al., 2020b; Thölke & De Fabritiis, 2022; Liao & Smidt, 2022; Batatia et al., 2022; Batzner et al., 2022), but, to the best of our knowledge, none of the existing methods can achieve *invariance*, *expressiveness*, and *scalability* at the same time as we show in Table 1.

To achieve an invariant model that is both expressive and scalable, we propose a simple but effective *Symmetrized GNN (SymGNN)*, which is E(3)-invariant in expectation. SymGNN achieves E(3) invariance by introducing a symmetrization module to aggregate embeddings produced under different orthogonal transformations of the graph structure. It is expressive as there is no higher-order information loss caused by "feature scalarization" as in models like SchNet (Schütt et al., 2017). For example, for the two graphs in Figure 2, SchNet will pass the same message to the central node of an isosceles triangle as if the middle node of a flat line, but SymGNN will easily distinguish between these two configurations (details in Appendix A.4). Also, SymGNN does not involve complex equivariant calculation, so it is much more scalable than methods relying on equivariant feature extraction (Liao & Smidt, 2022; Batatia et al., 2022; Batzner et al., 2022). In our experiments, we demonstrate that when applied to MG graphs, SymGNN outperforms a variety of widely used GNNs.

Methods	Invariance	Expressiveness	Scalability
GCN	✗	✓	✓
EGNN	?	✓	✓
SchNet	✓	✗	✓
MGCN	✓	✗	✓
FAENet	✓	✗	✓
DimeNet	✓	✓	✗
Torch-MD Net	✓	✓	✗
Equiformer	✓	✓	✗
SymGNN (ours)	E	✓	✓

Table 1. Comparison of different methods. ✓ means the model performs well evaluated by the corresponding category, ✗ means not, and ? means "unclear", or "possible after non-trivial extensions". The E indicates that the model satisfies the property in expectation.

Moreover, to better understand the EB prediction and benefit MG research, we also generate explanations along with the model prediction. Our proposed explanation method extends GNNExplainer (Ying et al., 2019) to the node regression task to generate edge-based structure explanations. It helps us to identify and visualize the importance of each edge in predicting an EB. We also show that the generated explanations match the medium-range order (MRO) hypothesis of MGs and possess unique topological properties that correlate with the optimal-volume cycles in a persistent diagram (Obayashi, 2018). Our findings provide further insights into the understanding of EBs, and our explanations can potentially benefit new scientific discoveries. We summarize our contributions as the following:

1. We formulate a material science research problem of predicting MG EBs as an ML problem of node regression on graphs.
2. We collect MG data for ML research, with precisely simulated EBs using ARTn.
3. We propose a simple but effective SymGNN model that exhibits E(3)-invariance in expectation and predicts MG EBs accurately and fast.
4. We generate explanations for EB predictions that match the MRO hypothesis, express unique topological properties, and provide insights for scientific discoveries.

2. Related work

2.1. ML in Material Science

The application of ML in materials science has seen significant advancements recently, with various models to tackle

different aspects of material science problems. Among these, GNNs have emerged as a powerful tool for representing and analyzing materials at the atomic level, owing to their ability to capture the complex relationships and interactions between atoms in a material. For example, estimating the propensity of individual atoms (Bapst et al., 2020), potential energy exhibited by a system of atoms (Schütt et al., 2017). In these settings, inductive bias of equivariance and invariance often plays a key role in the generalizability of network. For example, in our problem the EB only depends on local molecule configuration and thus are invariant on translation, rotation, and reflection of graphs. To incorporate this physical inductive bias, various invariant and equivariant GNNs have been proposed. Invariant GNNs often restrict graph features to be rotationally invariant, such as edge distances and angles, or reducing the inputs by projecting it onto PCA frames (Schütt et al., 2017; Gasteiger et al., 2020b;a; Duval et al., 2023b), whereas equivariant networks are proposed to leverage tensorial transformation that can extract equivariant node features (Schütt et al., 2021; Liao & Smidt, 2022; Batatia et al., 2022; Batzner et al., 2022; Thölke & De Fabritiis, 2022).

Several evaluation benchmarks for equivariant ML models on molecular dynamics (Bihani et al., 2024) and solid-state materials systems (Choudhary et al., 2020; Lee et al., 2023) have been proposed. EGRaffBench (Bihani et al., 2024) provides insights into the performance of various equivariant architectures in predicting forces in molecular systems, highlighting their potentials in simulating atomistic interactions. Moreover, JARVIS (Choudhary et al., 2020) and the MatSciML benchmark (Lee et al., 2023) represent significant efforts in benchmarking MLs to solid-state materials systems. These works demonstrate the potential of ML models, including various GNNs, in predicting properties and behaviors of solid-state materials, thereby aiding in materials design and discovery. Furthermore, a comprehensive overview of geometric GNNs for 3D atomic systems is provided by Duval et al. in their recent review (Duval et al., 2023a). This guide offers valuable insights into the development and application of GNNs in materials science, emphasizing the importance of geometric considerations in modeling atomic systems.

2.2. MGs and EBs

Understanding the relationship between the atomic structure and physical properties of MGs is one of the greatest challenges for both material science and condensed matter physics (Falk & Langer, 2011; Sun & Wang, 2015; Nicolas et al., 2018). However, the structure-property relationship of MGs is often challenging to characterize directly due to the complexity of the physical properties (Cubuk et al., 2017; Bapst et al., 2020). EBs describe the local roughness of the energy landscape by comparing the average energy difference around an atom's local neighbors. They are influential in MG dynamics and their physical properties (Berthier & Biroli, 2011; Kirchner et al., 2022), for example, the degree of ductility during fracture (Tang et al., 2021). Therefore, EBs can act as an important intermediary step when predicting the physical properties with the atomic structures as inputs (Debenedetti & Stillinger, 2001; Yu et al., 2012; Wang et al., 2020; Tang et al., 2021). ML methods have been applied to investigate the relationship between the atomic structures and physical properties in MG (Bapst et al., 2020). For EBs in particular, (Wang et al., 2020) explored using XGBoost to classify nodes with the highest 5 percent activation energy. Our work furthers the investigation of (Wang et al., 2020) by leveraging the natural graph structure using GNNs to perform a regression for EBs and generating insightful explanations.

3. Problem Setup and Preliminaries

3.1. EB Prediction with GNNs

The problem of predicting EBs of MGs can be formalized as a node regression problem on graphs. Under this formulation, atoms become nodes in a graph, and edges are constructed between nearby nodes. The MG data thus becomes a graph with n nodes and m edges. We represent the graph structure with G, which indicates all the edges and is normally represented in the form of an adjacency matrix. The node features are the atom types, which we represent with $\boldsymbol{Z} = \{\boldsymbol{z}^1, \boldsymbol{z}^2, \ldots, \boldsymbol{z}^n\}$. The edge features are the displacement vectors constructed from 3D node coordinates, which we represent with $\boldsymbol{X} = \{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^m \,|\boldsymbol{x}^i \in \mathbb{R}^3\}$. The regression task is to predict the EB label $y \in \mathbb{R}$ of each node with the graph structure and features as inputs, i.e., a model that maximizes $P(y|\, G, \boldsymbol{Z}, \boldsymbol{X})$. We further break down the prediction process into two steps. The first step encodes node and edge features to embeddings $\boldsymbol{H}$. The second step predicts y with G and $\boldsymbol{H}$ as inputs. The objective to maximize becomes the following,

$$P(y|\, G, \boldsymbol{Z}, \boldsymbol{X}) = \int_{\boldsymbol{H}} P(y|\, G, \boldsymbol{H}) P(\boldsymbol{H}|\, G, \boldsymbol{Z}, \boldsymbol{X}) d\boldsymbol{H} \tag{1}$$

We solve this problem with the state-of-the-art graph ML models - GNNs.

3.2. Orthogonality and Invariance

EB is invariant to Euclidean transformations of the atomic graph structure, for example, rotations, reflections, and translations, because it is the average energy needed for a node to hop between its current and nearby energy subbasins. Given that the graph is described with displacement vectors of relative positions, translations will be canceled, and the invariance to Euclidean transformations can be reduced to

the invariance to orthogonal transformations (Hall & Hall, 2013), which is defined as the following,

Definition 3.1 (Orthogonal Transformation)**.** A linear transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ is called an *orthogonal transformation* if it preserves the inner produce $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^d$, i.e., $\forall \boldsymbol{x}^1, \boldsymbol{x}^2 \in \mathbb{R}^d$, $\langle T(\boldsymbol{x}^1), T(\boldsymbol{x}^2) \rangle = \langle \boldsymbol{x}^1, \boldsymbol{x}^2 \rangle$. Then, the matrix form of T has $|\det(T)| = 1$. The *orthogonal group* in dimension d is the group of all such orthogonal transformations on $\mathbb{R}^d$ and is denoted as $\mathrm{O}(d)$.

We also state a well-known lemma in group theory (Hall & Hall, 2013) and a theorem by Euler (Slabaugh, 1999) for decomposing the orthogonal group and rotations respectively. They will be useful for modeling invariance.

Lemma 3.2 ($\mathrm{O}(3)$ Decomposition)**.** *The orthogonal group* $\mathrm{O}(3)$ *can be decomposed into rotations and non-rotations. The rotations also form a group denoted as* $\mathrm{SO}(3)$*, and it contains all transformations* R *whose matrix forms have* $\det(R) = 1$*. The non-rotations contain all the reflections and roto-reflections (also called improper rotation)* $\tilde{R}$*, whose matrix form have* $\det(\tilde{R}) = -1$*. Non-rotations can be denoted as* $P \cdot \mathrm{SO}(3)$*, with* P *being any reflection transformation through the origin.*

Theorem 3.3. *(Euler) Define the rotations around the three coordinate axes* x_1, x_2*, and* x_3 *in* $\mathbb{R}^3$ *by*

$$O_{x_1}(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

$$O_{x_2}(\beta) = \begin{bmatrix} \cos(\beta) & 0 & -\sin(\beta) \\ 0 & 1 & 0 \\ \sin(\beta) & 0 & \cos(\beta) \end{bmatrix}$$

$$O_{x_3}(\gamma) = \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then any rotation $R \in \mathrm{SO}(3)$ *can be written as* $R_{\alpha,\beta,\gamma} = O_{x_1}(\alpha) O_{x_2}(\beta) O_{x_3}(\gamma)$ *for some angles* $[\alpha, \beta, \gamma] \in [-\pi, \pi]^3$*. These angles are called the Euler angles.*

Then we formally introduce the invariant/equivariant transformation.

Definition 3.4 (Invariant/Equivariant Transformation)**.** Given a group K acts on $\mathbb{R}^d$. A transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ is invariant to K if $T(\boldsymbol{x}) = T(k \cdot \boldsymbol{x})$ and equivariant to K if $k \cdot T(\boldsymbol{x}) = T(k \cdot \boldsymbol{x})$ for all $k \in K$ and for all $\boldsymbol{x} \in \mathbb{R}^d$.

3.3. GNNExplainer

As a representative GNN explanation method, GNNExplainer seeks to explain GNN classifications by selecting an important edge-induced subgraph G_S that minimizes the entropy $H(\cdot)$ of the label Y. Since G_S is discrete, GNNExplainer learns a continuous distribution $\mathcal{G}$ over G_S that gives the minimal expected entropy, where $\mathcal{G}$ can be implemented as a learnable edge mask $\boldsymbol{M} \in \mathbb{R}^{|G|}$ applied on edges of G after a sigmoid function σ. Mathematically, the optimization objective is

$$\min_{\mathcal{G}} \mathbb{E}_{G_S \sim \mathcal{G}} H(Y|G = G_S) = \min_{\boldsymbol{M}} H(Y|G = \sigma(\boldsymbol{M}) \odot G) \tag{2}$$

4. Method

In this section, we present SymGNN for solving the EB prediction problem we formalized in Section 3.1. We first introduce the theory behind the core symmetrization module for capturing $\mathrm{O}(3)$ invariance in Section 4.1, then the full SymGNN model in Section 4.2, and finally how we apply explanation algorithms to SymGNN to reveal the connection between the atomic structures and EBs in Section 4.4.

4.1. Theory of Symmetrization Over $\mathrm{O}(3)$

Although EB is invariant to Euclidean transformations of the atomic graph structure, most GNNs are not designed to automatically capture such invariance. There are existing GNNs specialized for molecular graphs that can handle such invariance, but they either utilize scalarization that cannot handle higher-order information, or cannot scale up to graphs with thousands of nodes like MGs. We thus propose a symmetrization module that can better capture invariance and efficiently scale up. This section presents the theory behind the symmetrization.

For the node regression problem formalized in Section 3.1, $\boldsymbol{X}$ only represents one set of displacement vectors under one particular coordinate system. To achieve $\mathrm{O}(3)$-invariant (and thus $\mathrm{E}(3)$-invariant as explained in Section 3.2) predictions, we propose a symmetrization over all orthogonal transformations of $\boldsymbol{X}$, denoted as $\mathcal{X} = \{T(\boldsymbol{X}) \,|\, \forall T \in \mathrm{O}(3)\}$. Under symmetrization, we reformulate the feature encoding step in Equation 1, i.e.,$P(\boldsymbol{H} \,|\, G, \boldsymbol{Z}, \mathcal{X})$, as a probability integrated over $\mathcal{X}$, i.e.,

$$P(\boldsymbol{H} \,|\, G, \boldsymbol{Z}, \mathcal{X}) = \int_{T \in \mathrm{O}(3)} P(\boldsymbol{H} \,|\, G, \boldsymbol{Z}, T(\boldsymbol{X})) P(T) \, dT \tag{3}$$

Notice that a truly $\mathrm{O}(3)$-invariant model will give the same result for $P(\boldsymbol{H} \,|\, G, \boldsymbol{Z}, \boldsymbol{X})$ and $P(\boldsymbol{H} \,|\, G, \boldsymbol{Z}, \mathcal{X})$. In the new formulation, when maximizing $P(\boldsymbol{H} \,|\, G, \boldsymbol{Z}, \mathcal{X})$, the model will learn the desired invariance by foreseeing and aggregating different transformed graphs. To model such an integral, we first define the distribution of T on $\mathrm{O}(3)$ through the following two lemmas.

Lemma 4.1. *Any non-rotation* $\tilde{R} \in P \cdot \mathrm{SO}(3)$ *can be written as* $\tilde{R}_{\alpha,\beta,\gamma} = -O_{x_1}(\alpha) O_{x_2}(\beta) O_{x_3}(\gamma)$ *for some parameters* $[\alpha, \beta, \gamma] \in [-\pi, \pi]^3$.

Proof. Please refer to Appendix A.2. □

Intuitively, Theorem 3.3 says that any rotation in 3D can be decomposed into a combination of rotations that rotate only around the x_1-axis, x_2-axis, and x_3-axis and parameterized with the Euler angels. Similarly, Lemma 4.1 says a similar decomposition and parameterization can be achieved for non-rotations as well. Bring these two results together gives the following lemma for decomposing any orthogonal transformation $T \in \mathrm{O}(3)$.

Lemma 4.2. *Any orthogonal transformation $T \in \mathrm{O}(3)$ can be written as $T_{\lambda,\alpha,\beta,\gamma} = (-1)^{\lambda} O_{x_1}(\alpha) O_{x_2}(\beta) O_{x_3}(\gamma)$ for some parameters $\lambda \in \{0, 1\}$ and $[\alpha, \beta, \gamma] \in [-\pi, \pi]^3$.*

Proof. Follow from Lemma 3.2, Theorem 3.3, and Lemma 4.1. □

Lemma 4.2 allows the integral in Equation 3 to be reduced into an integral over λ and $[\alpha, \beta, \gamma]$ in Equation 4, which is the objective our GNN will model.

$$P(\boldsymbol{H} \mid G, \boldsymbol{Z}, \mathcal{X}) = \int_{\lambda,\alpha,\beta,\gamma} P(\boldsymbol{H} \mid G, \boldsymbol{Z}, T_{\lambda,\alpha,\beta,\gamma}(\boldsymbol{X})) P(T_{\lambda,\alpha,\beta,\gamma})\, d\lambda d\alpha d\beta d\gamma \tag{4}$$

4.2. Symmetrized GNN

We now present the full SymGNN model with an illustration shown in Figure 3. SymGNN consists of two sub-modules. The first is the symmetrization module mentioned above for producing $\mathrm{O}(3)$-invariant embeddings $\boldsymbol{H}$, which we indicate with $\boldsymbol{H} = Sym(G, \boldsymbol{Z}, \boldsymbol{X})$. The second is a prediction module that takes the symmetrized $\boldsymbol{H}$ and G to perform message passing with attention and then node regression.

The Sym module produces embeddings following the objective in Equation 4 with a learnable encoder Enc, i.e.,

$$\begin{aligned} \boldsymbol{H} &= Sym(G, \boldsymbol{Z}, \boldsymbol{X}) \\ &= \int_{\lambda,\alpha,\beta,\gamma} Enc(G, \boldsymbol{Z}, T_{\lambda,\alpha,\beta,\gamma}(\boldsymbol{X})) P(T_{\lambda,\alpha,\beta,\gamma}) d\lambda d\alpha d\beta d\gamma \end{aligned} \tag{5}$$

However, one challenge is that there are infinitely many $T \in \mathrm{O}(3)$, which makes the integral intractable. To model such an integral, we generate transformations $T_1, \ldots, T_k$ from $\mathrm{O}(3)$ by sampling λ and $[\alpha, \beta, \gamma]$ to approximate the Sym in Equation 5, which gives the $Sym_{T_1,\ldots,T_k}$ we use in practice.

$$Sym_{T_1,\ldots,T_k}(G, \boldsymbol{Z}, \boldsymbol{X}) = \frac{1}{k} \sum_{i=1}^{k} Enc(G, \boldsymbol{Z}, T_i(\boldsymbol{X})) \tag{6}$$

We show that $Sym_{T_1,\ldots,T_k}$ is $\mathrm{O}(3)$-invariant in expectation under assumptions of uniform distributions.

Theorem 4.3. *Assume $T_1, \ldots, T_k$ are random transformations that follow a uniform distribution over all $T \in \mathrm{O}(3)$. Then, Sym is $\mathrm{O}(3)$-invariant in expectation in the sense that $\mathbb{E}_{T_1,\ldots,T_k}[Sym_{T_1,\ldots,T_k}(G, \boldsymbol{Z}, \boldsymbol{X})] = \mathbb{E}_{T_1,\ldots,T_k}[Sym_{T_1,\ldots,T_k}(G, \boldsymbol{Z}, T_0\boldsymbol{X})]$ for any $T_0 \in \mathrm{O}(3)$.*

Proof. Please refer to Appendix A.3. □

Sym can learn from a variety of orthogonal transformations and achieve invariance. In practice, we fix λ to be $\mathrm{Bern}(0.5)$ to balance rotations and non-rotations uniformly, but we parametrize α, β, γ with learnable von Mises (Tikhonov) distributions (Mardia & Zemroch, 1975) instead of uniform. Learnable distributions help Sym more efficiently sample orthogonal transformation that benefits the prediction. The von Mises parameterization indeed leads to better empirical performance than uniform, since these distributions closely approximate the wrapped normal distribution on $[-\pi, \pi]$.

The second prediction module takes the invariant embeddings $\boldsymbol{H}$ produced by Sym to perform message passing and predict y. Given the complexity of the prediction problem and to enhance model expressiveness, we also compute attention of edge features and add skip connections during message passing. Specifically, we build up on the Edge Graph Attention Network (EGAT) (Kamiński et al., 2021) model to add edge features to the attention calculation in addition to the regular GAT.

Specifically, after the message from each node is computed, we first calculate an attention score a_{ij} over the edge between nodes i and j. Then the representation of node i in the $l+1$-th layer (e_i^{l+1}) is constructed as the attention-weighted average of the neighbor representations from the l-th layer. We show the formula in the following, where σ represents the non-linear activation function and $\mathcal{N}(i)$ represents the set of neighbors of node i.

$$a_{ij}^l = \frac{\exp x_{ij}^l}{\sum_{k \in \mathcal{N}(i)} \exp x_{ik}^l}, \quad e_i^{l+1} = \sigma\Big(\sum_{j \in \mathcal{N}(i)} a_{ij} e_j^l\Big)$$

4.3. Computation Time Analysis

We provide a theoretical analysis of the time complexity of SymGNN against three other baselines including SchNet, DimeNet, and Equiformer. For SymGNN, the time complexity is $O(kn + nd^2)$, where k denotes the number of sampled orthogonal transformation, n denotes the number of nodes, and d denotes the average number of neighbors a node has. The first kn term comes from the symmetrization of k orthogonal transformations, the second nd^2 term comes from the attention operation among a node's neighbors. Add these two terms together we essentially get $O(nd^2)$ (empirically we set $k = 6$). In comparison, SchNet is faster

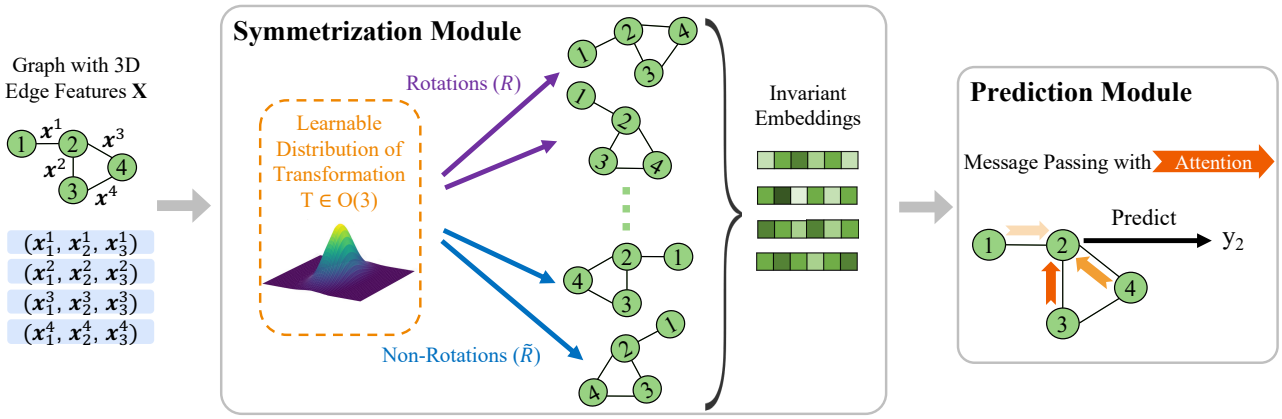

Figure 3. Illustration of the SymGNN framework. Given an input graph with node features being atom types and edge feature the relative distance, the symmetrization module of SymGNN aggregates encoding results on various orthogonally transformed graphs sampled from a learnable distribution to achieve $\mathrm{O}(3)$-invariant in expectation. The invariant embeddings are then passed to message-passing layers with attention to aggregate information and predict label.

Table 2. Theoretical Time Complexity.

	SymGNN	SchNet	Equiformer	DimeNet
Complexity	$O(kn + nd^2)$	$O(nd)$	$\Omega(nd^2)$	$O(n^4)$

as it is $O(nd)$. But for DimeNet, since it considers pairwise edge interaction, the time complexity grows at least as $O(m^2) = O(n^4)$, which makes it prohibitively slow for our graphs with thousands of nodes. For Equiformer, its exact time complexity is unknown to us. It is a transformer-based method where they change attention to equivariant attention and linear layer to equivariant linear layer using complex tensor operations, which implies that its big-O time complexity is lower bounded by $O(nd^2)$. Empirically, we found that training Equiformer is excessively slow for our large-scale graphs. A list of analyzed time complexity can be found in Table 2.

4.4. Explanations for Structure-EB Relationship

ML models have emerged as powerful tools in scientific research, and their utility can extend beyond mere predictions to explanations. This explanatory aspect is crucial because it aligns with the fundamental objective of ML for science: identifying patterns that can elude human analysis and understanding the underlying mechanisms that govern phenomena.

To make the best use of the SymGNN model and truly bolster the scientific research of MGs, we generate explanations to better reveal the structure-EB relationship. We choose GNNExplainer as a starting point for selecting a subgraph G_S with important edges. Since GNNExplainer was developed for classification problems, the cross-entropy-based objective does not apply to the regression problem of EB prediction. Therefore, we still learn an edge mask $\boldsymbol{M}$ on all edges, but modify the objective in Equation 2 by replacing entropy with mean squared error (MSE) as below, with f representing the SymGNN model.

$$\min_{\mathcal{G}} \mathbb{E}_{G_S \sim \mathcal{G}} \mathrm{MSE}(f(G_S)) = \min_{\boldsymbol{M}} \mathrm{MSE}(f(\sigma(\boldsymbol{M}) \odot G)) \tag{7}$$

This regression explainer considers all edges involved in the prediction of EB for one node and assigns a score to each edge. These scores represent the importance of their corresponding edges for making the prediction. In Section 6, we demonstrate that the important edges identified by our explainer match the MRO insights mentioned in previous material research and possess unique topological properties.

5. Experiments

We conduct experiments by first constructing an MG dataset with energy barriers simulated by molecular dynamics. Then we apply SymGNN to this dataset and compare its performance with other baseline models. We also perform ablation studies of the symmetrization module to show its effectiveness.

5.1. Dataset

The proposed Cu64Zr36 dataset. We employ molecular dynamics to simulate the behavior of a representative Cu64Zr36 MG subjected to shear deformation. The simulated MG system comprises 8000 atoms, generated through the conventional melting-quenching procedure. To evaluate

the influence of system size, we also simulate small systems with 3000 atoms. To ensure the statistical robustness of our findings, 9 independent metallic glass samples are generated. To obtain the energy barriers of atoms, we employ the activation-relaxation technique nouveau (ARTn) (Barkema & Mousseau, 1996; Cances et al., 2009) to calculate the energy barriers. The simulated results are used to construct a dataset consisting of nine graphs. Among them, six graphs are used for training, one graph is for validation, and two graphs are for testing. Each training/validation graph has 8,000 nodes and roughly 260,000 edges, and each test graph has 3,000 nodes and roughly 100,000 edges. In Appendix A.5, we provide a detailed description of the dataset construction process including the simulation temperature, pressure control, cooling rates, ARTn saddle point search algorithms, edge construction threshold, and etc.

Cu-Zr MGs from (Wang et al., 2020) We also tested our method on other metallic glass dataset. We adopted the dataset proposed in one of the previous work (Wang et al., 2020), which includes two additional Cu-Zr type metallic glasses Cu80Zr80 and Cu50Zr50. For each type of material the dataset contains two graphs each with 5000 nodes and around 650000 edges. We picked one as training graph and the other one for testing.

5.2. Experiment Settings

Baselines: We evaluate our model against a variety of other ML models including Graph Convolutional Network (GCN) (Kipf & Welling, 2017) withe edge features, E(n) Equivariant GNN (EGNN) (Satorras et al., 2022) that are designed to handle equivariant features, a non-graph based multi-layer perceptron (MLP) model, and various invariant baselines that are proposed to handle molecular data including SchNet (Schütt et al., 2017), MGCN (Lu et al., 2019), DimeNet (Gasteiger et al., 2020b), Torch-MD Net (Thölke & De Fabritiis, 2022), Equiformer (Liao & Smidt, 2022), and FAENet (Duval et al., 2023b),. Furthermore, we perform two ablation studies named SymGNN w/o symmetrization where we remove the symmtrization layer and Data Augmentation where we only aggregated the embedding from three fixed orthogonal transformation instead of a learned one. In addition we have include a simple baseline in which we use the absolute length of edge instead of its 3D coordinates as an input edge feature to achieve invariance. We also compared to MD based local sampling approximation that is widely used by material scientists (Krishnan et al., 2017; Sastry et al., 1998).

Evaluation: The predicted energy barriers are evaluated by the Pearson product-moment correlation coefficient against the true values following from previous work in material science literature (Bapst et al., 2020). We run each experiment 4 times with different random initializations. On our Cu64Zr36 dataset, we use the validation set to determine the best model and compute the score with the best model on the test set. For the other Cu-Zr MGs dataset, we compute the test accuracy on the final epoch since there is no validation set.

Implementation: For our proposed Cu64Zr36 dataset, we train 4-layer GNNs for 20,000 epochs using an Amsgrad optimizer (Reddi et al., 2019) with a learning rate of 0.0001. We adopt an early stopping scheme if the model's prediction score on the validation set did not improve for 1000 epochs. For the other Cu-Zr MGs dataset, we train each model to a fixed number of epochs as there is no validation set. For those models that have smaller scale and faster convergence, i.e., MLP, GCN, EGNN, EGAT, SchNet, and MGCN, we train to 5000 epochs. For SymGNN, we trained the model to 10000 epochs for better convergence. In all the datasets for SymGNN, the distribution over the angles α, β, and γ is parameterized by the von Mises (Tikhonov) distribution.

5.3. Prediction Results

We report the results of SymGNN and other baselines in Table 3. It can be seen from the table that SymGNN outperforms the baselines by a large amount and exhibits a much stronger generalization power. When we remove the symmetrization module, (i.e. SymGNN w/o symmetrization), the ablated model cannot generalize well, and a similar performance drops is observed when we only aggregates embedding from three fixed orthogonal transformations. This demonstrates the effectiveness of the symmetrization module. Also we observe that models capable of handling invariance can lead to much better result compared to the ones that cannot, which again highlights the importance of symmetrization module in achieving good prediction performance. In addition, we also ran Equiformer, Torch-MD Net, and DimeNet as our baselines. However, we noticed that the training time for these methods are prohibitively long (i.e longer than 2 days) on our dataset.

5.4. Computation Time Analysis

We notice that SymGNN reaches high performance without dramatically increase both the training cost or the inference cost. Table 4 provides an empirical time comparison for the time needed to train the model for one epoch. We observe that DimeNet would run indefinitely for our larger graph, and the time taken by Equiformer is also prohibitively long. Table 5 shows a inference time comparison. Compared to traditional MD simulation, our ML-based approach needs much fewer computation resources and is much more efficient. For precise MD simulation with ARTn, the calculation of the energy barrier for each atom takes around 20 minutes in a supercomputer with 16 parallel threads. Therefore, for a MG system that has the size of our test graph, i.e.,

Table 3. Training and testing scores of the molecular dynamics (MD) method and machine learning (ML) methods. Test results are with the best model on the validation set. Our SymGNN significantly outperforms the MD method and achieves the best among all the ML methods.

	Methods	Cu64Zr36	Cu80Zr20	Cu50Zr50
MD	Local Sampling (Sastry et al., 1998)	0.3614	–	–
Non-Invariant ML	MLP	0.0575 ± 0.0127	0.0727 ± 0.0154	-0.0652 ± 0.0099
	GCN with Edge Features	0.5123 ± 0.0507	0.2478 ± 0.0051	0.1395 ± 0.0068
Invariant ML	E(n) Equivariant GNN	0.2588 ± 0.0077	0.1382 ± 0.0113	0.1381 ± 0.0098
	EGAT (Edge Length as 1D Feature)	0.7264 ± 0.0063	0.5489 ± 0.0218	0.1571 ± 0.0095
	SchNet	0.7588 ± 0.0088	0.2505 ± 0.0128	0.1808 ± 0.0106
	MGCN	0.7352 ± 0.0066	0.1793 ± 0.0133	0.1596 ± 0.0033
	FAENet	0.6603 ± 0.0218	0.2947 ± 0.0171	0.2214 ± 0.0160
Ours	SymGNN	$\mathbf{0.7859 \pm 0.0056}$	$\mathbf{0.6084 \pm 0.0167}$	$\mathbf{0.5862 \pm 0.0277}$
	SymGNN w/o symmetrization	0.2669 ± 0.0371	0.2283 ± 0.0256	0.1135 ± 0.0129
	Data Augmentation	0.6614 ± 0.0285	0.3304 ± 0.0201	0.2135 ± 0.0337

Table 4. Training time comparison for one epoch.

	SymGNN	SchNet	Equiformer	DimeNet
Time	3 secs	1 sec	82 mins	-

Table 5. Inference time comparison on an MG with 3,000 atoms.

	SymGNN	ARTn	MD local sampling
Time	0.26 seconds	41 days	150 mins

3,000 atoms, the computation will take $\frac{20\times3000}{60\times24} \approx 41$ days. Even for the much faster and less inaccurate local sampling method, the inference time for this MG system can take 150 minutes. In contrast, SymGNN's inference time on the test graph is almost negligible.

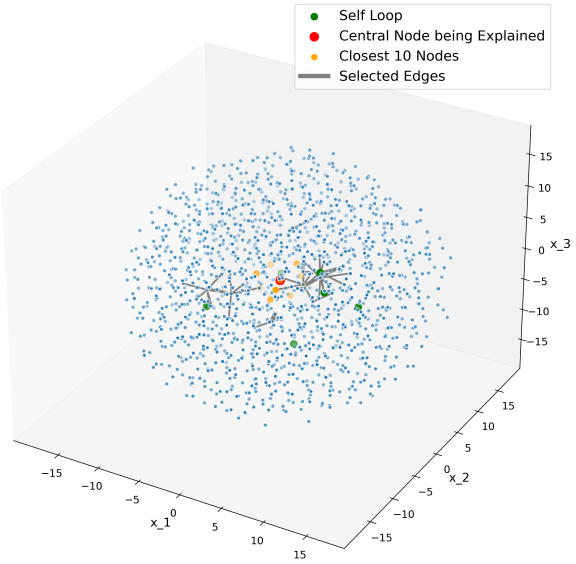

Figure 4. Global explanation visualization. We notice that many of selected edges are of 2 hop neighborhood from the central node.

6. Explanation and Analysis

We produce explanation using the method in Section 4.4. We first provide visualizations to qualitatively show our explanation, and then we do a quantitative evaluation by connecting of our explanation to the MRO hypothesis and the topological data analysis (TDA) to reveal more insights.

6.1. Explanation Visualization and MRO

We visualize our explanation for a randomly sampled node. We provide both the global and the local version. All atoms are plotted in their actual 3D coordinates. The global version of the explanation is presented in Figure 4, where we visualize the top 50 important edges. The local version is in Appendix Figure 6, where we zoom in to the top 10 closet nodes. From the visualizations, we see that edges close to the central node or far away from the central node may both be selected. Also, as we can see from the local version of the explanation, few edges within the top 10 closet edges is selected by our explanation.

Material scientists proposed an MRO hypothesis, which basically says the atoms that lie in a range of medium distance (5 - 10 Å) from the central atom play a more important role in determining its MG properties (Ma et al., 2009; Sheng

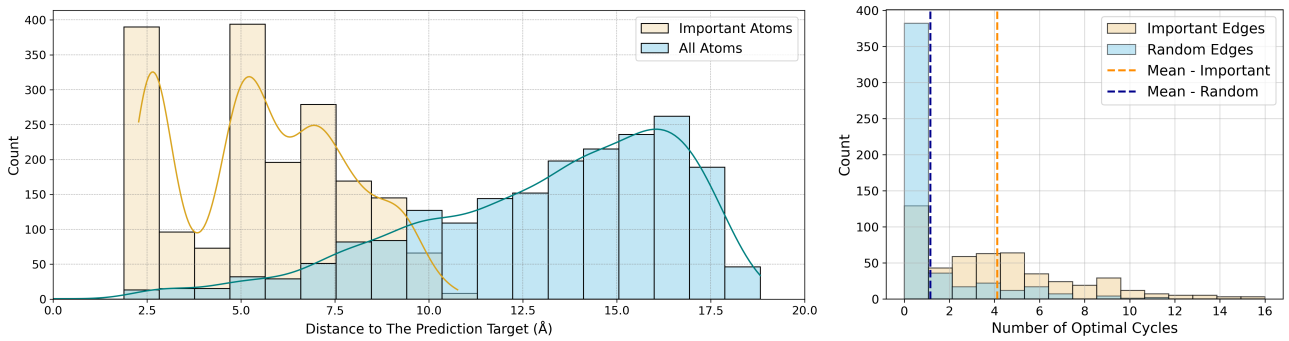

Figure 5. (a) Distribution of distance to the prediction target (central node) of important atoms identified by our explanation vs. all atoms. (b)Distribution of the number of cycles involved in important edges identified in our explanation vs. randomly selected edges.

et al., 2006; Nomoto et al., 2021; Egami et al., 2023). We aggregate our selected edges to nodes (atoms) and plot the atom importance against their distance to the central node in Figure 5 (a). We see there are two modes of the important atoms, one for the closest ones and one for the medium-range ones, which matches the MRO hypothesis.

6.2. Edge Importance vs. TDA

We perform TDA to further understand our edge importance explanations and see if the results recover meaningful topological structures.

Persistent Homology (PH) PH is a widely used TDA method, a field of study that applies concepts from algebraic topology to data analysis (Barannikov, 1994; Zomorodian & Carlsson, 2004; Edelsbrunner et al., 2008). PH examines how topological features such as connected components, holes, or voids, emerge and disappear as one moves through different scales in a dataset. The persistence of certain topological features across scales can reveal important insights about the underlying structure of the data, making PH a powerful tool for material science. This method has also been applied to metallic glass to uncover important topological structures (Sørensen et al., 2020). In PH, the concepts of "birth" and "death" are the essential quantities we would like to study, which visually represent the lifespan of topological features in a dataset. "Birth" refers to the scale at which a feature, like a connected component or a hole, first appears during the filtration process, while "death" denotes the scale at which this feature disappears or merges.

Explanation Results In our case, we apply PH to study the emergence and death of 1D hole as we increase the radius of a ball surrounding each atom. We perform the inverse analysis to pair the hole with a representative optimal cycle (Obayashi, 2018). In this way, each edge in the graph can be associated with a sequence of births and deaths of the cycles that it has participated in. We perform statistical analysis to see if there is significant difference between selected edges by our explanation and other edges. We plot and compare the distribution of the number of optimal cycles involved in the highest importance edges selected by our explanation versus randomly selected edges over multiple central nodes that are being explained in Figure 5 (b). We found that on average the importance edges participates in much more cycles compared to others, and there is a clear trend in the decrease in cycle number as the importance of edges decrease. The mean of the number of cycles involved in by edges in the four different group can be found in Table 6.

Table 6. Average number of optimal cycles participated by edges with high, medium, and low importance according to the explanation and randomly selected edges.

Edge Importance	High	Medium	Low	Random
Avg # Optimal Cycles	4.130	1.202	0.874	1.148

7. Conclusion

In this paper, we study the connection between the local atomic structures of MGs and their EBs of the energy landscape. We formalize this problem as node regression on graphs and propose SymGNN to solve the problem by effectively capturing the invariance of orthogonal transformations of the graph. We compare SymGNN with several baseline models and demonstrate that SymGNN performs the best. In addition, we extend the GNNExplainer to regression tasks and generate explanations. We further investigate the explanations with MRO and PH. We show a strong correlation between the importance of edge and the number of optimal cycles they involved in. Our work enables effective prediction and interpretation of MG EBs, bolstering material science research.

Acknowledgements

This work was partially supported by NSF 2211557, NSF 1937599, NSF 2119643, NSF 2303037, NSF 2312501, NASA, SRC JUMP 2.0 Center, Amazon Research Awards, and Snapchat Gifts.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning, especially on the area AI for material science. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Bansal, N. P. and Doremus, R. H. *Handbook of glass properties*. Elsevier, 2013.

Bapst, V., Keck, T., Grabska-Barwińska, A., Donner, C., Cubuk, E. D., Schoenholz, S. S., Obika, A., Nelson, A. W. R., Back, T., Hassabis, D., and Kohli, P. Unveiling the predictive power of static structure in glassy systems. *Nature Physics*, 16(4):448–454, April 2020. doi: 10.1038/s41567-020-0842-8.

Barannikov, S. The framed morse complex and its invariants. *Advances in Soviet Mathematics*, 21:93–116, 1994.

Barkema, G. and Mousseau, N. Event-based relaxation of continuous disordered systems. *Physical review letters*, 77(21):4358, 1996.

Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems*, 35: 11423–11436, 2022.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022.

Berthier, L. and Biroli, G. Theoretical perspective on the glass transition and amorphous materials. *Reviews of modern physics*, 83(2):587, 2011.

Bihani, V., Mannan, S., Pratiush, U., Du, T., Chen, Z., Miret, S., Micoulaut, M., Smedskjaer, M. M., Ranu, S., and Krishnan, N. A. Egraffbench: evaluation of equivariant graph neural network force fields for atomistic simulations. *Digital Discovery*, 3(4):759–768, 2024.

Billingsley, P. *Probability and measure*. John Wiley & Sons, 2017.

Cances, E., Legoll, F., Marinica, M.-C., Minoukadeh, K., and Willaime, F. Some improvements of the activation-relaxation technique method for finding transition pathways on potential energy surfaces. *The Journal of chemical physics*, 130(11), 2009.

Cao, Y., Li, J., Kou, B., Xia, C., Li, Z., Chen, R., Xie, H., Xiao, T., Kob, W., Hong, L., et al. Structural and topological nature of plasticity in sheared granular materials. *Nature communications*, 9(1):2911, 2018.

Cheng, Y. and Ma, E. Atomic-level structure and structure–property relationship in metallic glasses. *Progress in materials science*, 56(4):379–473, 2011.

Choudhary, K., Garrity, K. F., Reid, A. C., DeCost, B., Biacchi, A. J., Hight Walker, A. R., Trautt, Z., Hattrick-Simpers, J., Kusne, A. G., Centrone, A., et al. The joint automated repository for various integrated simulations (jarvis) for data-driven materials design. *npj computational materials*, 6(1):173, 2020.

Cubuk, E. D., Ivancic, R., Schoenholz, S. S., Strickland, D., Basu, A., Davidson, Z., Fontaine, J., Hor, J. L., Huang, Y.-R., Jiang, Y., et al. Structure-property relationships from universal signatures of plasticity in disordered solids. *Science*, 358(6366):1033–1037, 2017.

Debenedetti, P. G. and Stillinger, F. H. Supercooled liquids and the glass transition. *Nature*, 410(6825):259–267, 2001.

Ding, J., Patinet, S., Falk, M. L., Cheng, Y., and Ma, E. Soft spots and their structural signature in a metallic glass. *Proceedings of the National Academy of Sciences*, 111 (39):14052–14056, 2014.

Duval, A., Mathis, S. V., Joshi, C. K., Schmidt, V., Miret, S., Malliaros, F. D., Cohen, T., Liò, P., Bengio, Y., and Bronstein, M. A hitchhiker's guide to geometric gnns for 3d atomic systems. *arXiv preprint arXiv:2312.07511*, 2023a.

Duval, A. A., Schmidt, V., Hernández-Garcıa, A., Miret, S., Malliaros, F. D., Bengio, Y., and Rolnick, D. Faenet: Frame averaging equivariant gnn for materials modeling. In *International Conference on Machine Learning*, pp. 9013–9033. PMLR, 2023b.

Edelsbrunner, H., Harer, J., et al. Persistent homology-a survey. *Contemporary mathematics*, 453(26):257–282, 2008.

Egami, T., Dmowski, W., and Ryu, C. W. Medium-range order resists deformation in metallic liquids and glasses. *Metals*, 13(3):442, 2023.

Falk, M. L. and Langer, J. S. Deformation and failure of amorphous, solidlike materials. *Annu. Rev. Condens. Matter Phys.*, 2(1):353–373, 2011.

Fan, Y., Iwashita, T., and Egami, T. How thermally activated deformation starts in metallic glass. *Nature communications*, 5(1):5083, 2014.

Fan, Y., Iwashita, T., and Egami, T. Energy landscape-driven non-equilibrium evolution of inherent structure in disordered material. *Nature communications*, 8(1):15417, 2017.

Gasteiger, J., Giri, S., Margraf, J. T., and Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020a.

Gasteiger, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020b.

Hall, B. C. and Hall, B. C. *Lie groups, Lie algebras, and representations*. Springer, 2013.

Hiraoka, Y., Nakamura, T., Hirata, A., Escolar, E. G., Matsue, K., and Nishiura, Y. Hierarchical structures of amorphous solids characterized by persistent homology. *Proceedings of the National Academy of Sciences*, 113(26): 7035–7040, 2016.

Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, 31(3):1695, 1985.

Jay, A., Gunde, M., Salles, N., Poberžnik, M., Martin-Samos, L., Richard, N., de Gironcoli, S., Mousseau, N., and Hémeryck, A. Activation–relaxation technique: An efficient way to find minima and saddle points of potential energy surfaces. *Computational Materials Science*, 209:111363, 2022.

Kamiński, K., Ludwiczak, J., Jasiński, M., Bukala, A., Madaj, R., Szczepaniak, K., and Dunin-Horkawicz, S. Rossmann-toolbox: a deep learning-based protocol for the prediction and design of cofactor specificity in Rossmann fold proteins. *Briefings in Bioinformatics*, 23(1):bbab371, 09 2021. ISSN 1477-4054. doi: 10.1093/bib/bbab371. URL https://doi.org/10.1093/bib/bbab371.

Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks, 2017.

Kirchner, K. A., Cassar, D. R., Zanotto, E. D., Ono, M., Kim, S. H., Doss, K., Bødker, M. L., Smedskjaer, M. M., Kohara, S., Tang, L., et al. Beyond the average: spatial and temporal fluctuations in oxide glass-forming systems. *Chemical reviews*, 123(4):1774–1840, 2022.

Krishnan, N. A., Wang, B., Yu, Y., Le Pape, Y., Sant, G., and Bauchy, M. Enthalpy landscape dictates the irradiation-induced disordering of quartz. *Physical Review X*, 7(3): 031019, 2017.

Lee, K. L. K., Gonzales, C., Nassar, M., Spellings, M., Galkin, M., and Miret, S. Matsciml: A broad, multi-task benchmark for solid-state materials modeling. *arXiv preprint arXiv:2309.05934*, 2023.

Liao, Y.-L. and Smidt, T. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

Liu, Z., Li, R., Liu, G., Su, W., Wang, H., Li, Y., Shi, M., Luo, X., Wu, G., and Zhang, T. Microstructural tailoring and improvement of mechanical properties in cuzr-based bulk metallic glass composites. *Acta Materialia*, 60(6-7): 3128–3139, 2012.

Lu, C., Liu, Q., Wang, C., Huang, Z., Lin, P., and He, L. Molecular property prediction: A multilevel quantum interactions modeling perspective. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 1052–1060, 2019.

Ma, D., Stoica, A. D., and Wang, X.-L. Power-law scaling and fractal nature of medium-range order in metallic glasses. *Nature materials*, 8(1):30–34, 2009.

Mardia, K. and Zemroch, P. Algorithm as 86: The von mises distribution function. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 24(2):268–272, 1975.

Mendelev, M., Kramer, M., Ott, R., Sordelet, D., Yagodin, D., and Popel, P. Development of suitable interatomic potentials for simulation of liquid and amorphous cu–zr alloys. *Philosophical Magazine*, 89(11):967–987, 2009.

Mousseau, N., Béland, L. K., Brommer, P., Joly, J.-F., El-Mellouhi, F., Machado-Charry, E., Marinica, M.-C., and Pochet, P. The activation-relaxation technique: Art nouveau and kinetic art. *Journal of Atomic and Molecular Physics*, 2012, 2012.

Nicolas, A., Ferrero, E. E., Martens, K., and Barrat, J.-L. Deformation and flow of amorphous solids: Insights from elastoplastic models. *Reviews of Modern Physics*, 90(4): 045006, 2018.

Nomoto, K., Ceguerra, A. V., Gammer, C., Li, B., Bilal, H., Hohenwarter, A., Gludovatz, B., Eckert, J., Ringer, S. P., and Kruzic, J. J. Medium-range order dictates local hardness in bulk metallic glasses

. *Materials today*, 44:48–57, 2021.

Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics*, 81(1):511–519, 1984.

Obayashi, I. Volume-optimal cycle: Tightest representative cycle of a generator in persistent homology. *SIAM Journal on Applied Algebra and Geometry*, 2(4):508–534, 2018.

Patinet, S., Vandembroucq, D., and Falk, M. L. Connecting local yield stresses with plastic activity in amorphous solids. *Physical review letters*, 117(4):045501, 2016.

Pauly, S., Gorantla, S., Wang, G., Kühn, U., and Eckert, J. Transformation-mediated ductility in cuzr-based bulk metallic glasses. *Nature materials*, 9(6):473–477, 2010.

Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, 117(1): 1–19, 1995.

Reddi, S. J., Kale, S., and Kumar, S. On the convergence of adam and beyond. *arXiv preprint arXiv:1904.09237*, 2019.

Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., van Hoesel, C., Schopmans, H., Sommer, T., et al. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1): 93, 2022.

Sastry, S., Debenedetti, P. G., and Stillinger, F. H. Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid. *Nature*, 393(6685):554–557, 1998.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks, 2022.

Schroers, J., Hodges, T. M., Kumar, G., Raman, H., Barnes, A. J., Pham, Q., and Waniuk, T. A. Thermoplastic blow molding of metals. *Materials Today*, 14(1):14–19, 2011. ISSN 1369-7021. doi: https://doi.org/10.1016/S1369-7021(11)70018-9. URL https://www.sciencedirect.com/science/article/pii/S1369702111700189.

Schütt, K., Kindermans, P.-J., Sauceda Felix, H. E., Chmiela, S., Tkatchenko, A., and Müller, K.-R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.

Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Sheng, H., Luo, W., Alamgir, F., Bai, J., and Ma, E. Atomic packing and short-to-medium-range order in metallic glasses. *Nature*, 439(7075):419–425, 2006.

Slabaugh, G. G. Computing euler angles from a rotation matrix. *Retrieved on August*, 6(2000):39–63, 1999.

Sørensen, S. S., Biscio, C. A., Bauchy, M., Fajstrup, L., and Smedskjaer, M. M. Revealing hidden medium-range order in amorphous materials using topological data analysis. *Science Advances*, 6(37):eabc2320, 2020.

Starr, F. W., Sastry, S., Douglas, J. F., and Glotzer, S. C. What do we learn from the local geometry of glass-forming liquids? *Physical review letters*, 89(12):125501, 2002.

Sun, B. and Wang, W. The fracture of bulk metallic glasses. *Progress in Materials Science*, 74:211–307, 2015.

Tang, L., Ma, G., Liu, H., Zhou, W., and Bauchy, M. Bulk metallic glasses' response to oscillatory stress is governed by the topography of the energy landscape. *The Journal of Physical Chemistry B*, 124(49):11294–11298, 2020.

Tang, L., Liu, H., Ma, G., Du, T., Mousseau, N., Zhou, W., and Bauchy, M. The energy landscape governs ductility in disordered materials. *Materials Horizons*, 8(4):1242–1252, 2021.

Thölke, P. and De Fabritiis, G. Torchmd-net: Equivariant transformers for neural network based molecular potentials. *arXiv preprint arXiv:2202.02541*, 2022.

Trexler, M. M. and Thadhani, N. N. Mechanical properties of bulk metallic glasses. *Progress in Materials Science*, 55(8):759–839, 2010.

Wang, Q., Ding, J., and Ma, E. Predicting the propensity for thermally activated β events in metallic glasses via interpretable machine learning, 2020.

Xu, B., Falk, M. L., Li, J., and Kong, L. Predicting shear transformation events in metallic glasses. *Physical review letters*, 120(12):125503, 2018.

Ying, R., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. Gnnexplainer: Generating explanations for graph neural networks, 2019.

Yu, H.-B., Samwer, K., Wu, Y., and Wang, W. H. Correlation between β relaxation and self-diffusion of the smallest constituting atoms in metallic glasses. *Physical review letters*, 109(9):095508, 2012.

Zomorodian, A. and Carlsson, G. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pp. 347–356, 2004.

A. Appendix

A.1. The EB Prediction Problem

In this section, we present a succinct high-level motivation for our work. Material scientists aim to use atomic structures of materials to predict their properties, such as ductility. However, this direct prediction is a challenging task. An alternative approach is to use an easier-to-predict intermediate quantity, e.g. the energy barriers, as a stepping stone. In other words, there is a shift from the paradigm 1:

$$\textbf{structures} - predict \rightarrow \textbf{properties}$$

to paradigm 2:

$$\textbf{structures} - predict \rightarrow \textbf{energy barriers} - analyze \rightarrow \textbf{properties}$$

The second paradigm has shown promising results, which is the focus of this work.

A.2. Proof of Lemma 4.1

In this section, we prove the non-rotation decomposition Lemma 4.1 stated in Section 4.1.

Proof. From Lemma 3.2, we know any non-rotation $\tilde{R} \in P \cdot \mathrm{SO}(3)$ has $\det \tilde{R} = -1$. By linearity of $\tilde{R}$ we know that

$$\langle -\tilde{R}(\boldsymbol{x}^1), -\tilde{R}(\boldsymbol{x}^2)\rangle = \langle \tilde{R}(-\boldsymbol{x}^1), \tilde{R}(-\boldsymbol{x}^2)\rangle = \langle -\boldsymbol{x}^1, -\boldsymbol{x}^2\rangle = \langle \boldsymbol{x}^1, \boldsymbol{x}^2\rangle$$

which shows hat $-\tilde{R}$ is also orthogonal according to Definition 3.1. As we know that $-\tilde{R}$ will have $\det -\tilde{R} = (-1)^3 \cdot -1 = 1$, by Lemma 3.2 again we know that $-\tilde{R}$ is a rotation. By Euler Theorem 3.3, we know there exists $[\alpha, \beta, \gamma] \in [0, 2\pi]^3$ such that $-\tilde{R} = O_{x_1}(\alpha)O_{x_2}(\beta)O_{x_3}(\gamma)$, which implies that $\tilde{R} = -O_{x_1}(\alpha)O_{x_2}(\beta)O_{x_3}(\gamma)$. □

A.3. Proof of Theorem 4.3

In this section, we prove the invariance in expectation Theorem 4.3 stated in Section 4.2.

Theorem A.1. *Assume $T_1, \ldots, T_k$ are random transformations that follow a uniform distribution over all $T \in \mathrm{O}(3)$. Then, Sym is $\mathrm{O}(3)$-invariant in expectation in the sense that $\mathbb{E}_{T_1,\ldots,T_k}[Sym_{T_1,\ldots,T_k}(G, \boldsymbol{Z}, \boldsymbol{X})] = \mathbb{E}_{T_1,\ldots,T_k}[Sym_{T_1,\ldots,T_k}(G, \boldsymbol{Z}, T_0\boldsymbol{X})]$ for any $T_0 \in \mathrm{O}(3)$.*

Proof. From Lemma 4.2, we know that all $T \in \mathrm{O}(3)$ have the form of $T = (-1)^\lambda O_{x_1}(\alpha)O_{x_2}(\beta)O_{x_3}(\gamma)$ for $\lambda \in \{0, 1\}$ and $[\alpha, \beta, \gamma] \in [-\pi, \pi]^3$. A uniform distribution over all $T \in \mathrm{O}(3)$ implies $\lambda \sim \mathrm{Bern}(0.5)$ and $[\alpha, \beta, \gamma] \sim \mathrm{Unif}([-\pi, \pi]^3)$.

Now consider a specific orthogonal transformation $T_0 = (-1)^{\lambda_0} O_{x_1}(\alpha_0)O_{x_2}(\beta_0)O_{x_3}(\gamma_0)$. Then its composition with the uniformed distributed T is

$$T \circ T_0 = (-1)^{\lambda+\lambda_0} O_{x_1}(\alpha + \alpha_0)O_{x_2}(\beta + \beta_0)O_{x_3}(\gamma + \gamma_0)$$

By the Bernoulli assumption, we get $(-1)^\lambda \sim (-1)^{\lambda+\lambda_0}$ as they both follow a discrete distribution on $\{-1, 1\}$ with probability 0.5 of each value. Moreover, $\alpha \sim \mathrm{Unif}[-\pi, \pi]$ implies $\alpha + \alpha_0 \sim \mathrm{Unif}[-\pi + \alpha_0, \pi + \alpha_0]$. Since this only shifts the interval that supports the uniform distribution, the joint distribution of $(\cos(\alpha), \sin(\alpha)) \sim (\cos(\alpha + \alpha_0), \sin(\alpha + \alpha_0))$ due to periodicity, which further implies $O_{x_1}(\alpha) \sim O_{x_1}(\alpha + \alpha_0)$. Similarly for $O_{x_2}(\beta)$ and $O_{x_3}(\gamma)$. Given the random variables (matrices) $(-1)^\lambda, O_{x_1}(\alpha), O_{x_2}(\beta)$, and $O_{x_3}(\gamma)$ are independent, we concluded that $T \sim T \circ T_0$.

Now consider $Sym_{T_1,\ldots,T_k}(G, \boldsymbol{Z}, \boldsymbol{X}) = \frac{1}{k}\sum_{i=1}^k Enc(G, \boldsymbol{Z}, T_i(\boldsymbol{X}))$. For each i, $T_i \sim T_i \circ T_0$ implies $T_i(\boldsymbol{X}) \sim T_i \circ T_0(\boldsymbol{X})$, and thus $Enc(G, \boldsymbol{Z}, T_i(\boldsymbol{X})) \sim Enc(G, \boldsymbol{Z}, T_i \circ T_0(\boldsymbol{X}))$ by transformation of random variables (Billingsley, 2017). Therefore, we also get $\mathbb{E}_{T_i}[Enc(G, \boldsymbol{Z}, T_i(\boldsymbol{X}))] = \mathbb{E}_{T_i}[Enc(G, \boldsymbol{Z}, T_i \circ T_0(\boldsymbol{X}))]$. Finally, by linearity of expectation, $\mathbb{E}_{T_1,\ldots,T_k}[Sym_{T_1,\ldots,T_k}(G, \boldsymbol{Z}, \boldsymbol{X})] = \mathbb{E}_{T_1,\ldots,T_k}[Sym_{T_1,\ldots,T_k}(G, \boldsymbol{Z}, T_0\boldsymbol{X})]$. □

A.4. Analysis of expressiveness of SymGNN

In this section, we show by examples that SymGNN can capture more complex interactions between molecules compared to SchNet like methods. Consider two three molecule systems that have the following configurations where the atom type for node 2 is two whereas the atom type for node 1 and node 3 is one:

(a) 1: $(-2, 0, 0)$, 2: $(0, 0, 0)$, 3:$(2, 0, 0)$

(b) 1: $(-\sqrt{2}, \sqrt{2}, 0)$, 2: $(0, 0, 0)$, 3:$(\sqrt{2}, \sqrt{2}, 0)$

Notice that the edge distance between node 2 to each of node 1 and node 3 are two in both of these configurations, so SchNet cannot distinguish these two configurations based on the embedding of node 2, as it only takes into account the distance information. However, we shall see that SymGNN can differentiate these configurations based on node 2's embedding as it considers also higher-order information. To ease the computation, we consider a minimal setting of SymGNN where the encoder is the identity function, and each time two orthogonal transformations will be applied to the graph and then aggregated. Finally, node 2's embedding is calculated by simply summing up the message passed between node 2 and node 1 and between node 2 and node 3. Suppose the two orthogonal transformations sampled are a counterclockwise rotation in xy-plane by 45 degrees and a reflection around y-axis. It can be calculated that SymGNN will give the embedding $(0, 0, 0)$ for node two in the first configuration, but the second configuration gives $(2, 2 + 2\sqrt{2}, 0)$. Similar example can be given to show that SymGNN can also detect configurations that have equivalent angle structure, and thus we know SymGNN truly considers higher level information compared to those invariant methods that are based on scalerization.

A.5. A Detailed Dataset Construction Process

We employ molecular dynamics to simulate the behavior of a representative Cu64Zr36 metallic glass (MG) subjected to shear deformation. The simulated MG system comprises 8000 atoms, generated through the conventional melting-quenching procedure with varied cooling rates spanning from 10^{14} to 10^{10} K/s. To evaluate the influence of system size, we also simulate small system (i.e., 3000 atoms). To initiate the simulation, the sample is initially melted at 2000K under zero pressure for 1ns, facilitating the erasure of its initial configuration memory. Temperature and pressure control are maintained through the isothermal-isobaric (NPT) ensemble, employing a Nosé-Hoover thermostat (Nosé, 1984; Hoover, 1985). Subsequently, the liquified state is rapidly quenched to 1K, with cooling rates ranging from 10^{14} to 10^{10} K/s. The resulting glassy structure is further relaxed to its local energy minimum through energy minimization, utilizing the conjugate gradient algorithm. The interatomic interactions within the system are described using the embedded-atom method (EAM) potential (Mendelev et al., 2009). To ensure the statistical robustness of our findings, 9 independent metallic glass samples are generated for each cooling rate. A timestep of 1fs is adopted for all simulations, and the entire set of simulations is carried out using the LAMMPS package (Plimpton, 1995).

To obtain the energy barriers of atoms, we employ the activation-relaxation technique nouveau (ARTn) (Barkema & Mousseau, 1996; Cances et al., 2009) to calculate the energy barriers within MGs. Specifically, starting from a local energy minimum in the landscape, initial perturbations are introduced to a chosen atom and its nearest neighbors. This perturbation allows exploration along a direction of negative curvature, increasing the likelihood of locating a saddle point in the energy landscape. The Lanczos algorithm (Barkema & Mousseau, 1996) is then applied to guide the system to the saddle point by following the direction of negative curvature. A force tolerance of $0.05eV/\text{Å}^{-2}$ is chosen to ensure convergence of the saddle points. In accordance with previous investigations (Fan et al., 2014; 2017; Xu et al., 2018), 20 searches for saddle points are conducted for each atom. Consequently, the ARTn exploration focuses on determining the average energy barrier associated with atoms. This parameter is recognized as a key factor influencing the propensity for plastic rearrangement in disordered materials (Tang et al., 2020; 2021). The simulated raw dataset initially only contains nodes (atoms) along with their types and 3D coordinates. We construct edges between two nodes if their Euclidean distance is smaller than a threshold, which is chosen to be $5\text{Å} = 10^{-10}m$.

A.6. CuZr-Based MGs as Representative Examples

In this section, we discuss choice of focusing on Cu-Zr based MGs in our dataset. First, Cu-Zr-based metallic glass is one of the most widely investigated MGs due to its outstanding mechanical properties and good glass-forming ability (Cheng & Ma, 2011). Many well-known studies on MGs, such as those focusing on mechanical properties and ductility (Liu et al., 2012; Pauly et al., 2010), are centered on Cu-Zr. Additionally, Cu-Zr has been used as a standard MG example in ML research to study β processes (Wang et al., 2020) and perform hierarchical structure analysis (Hiraoka et al., 2016). Cu64Zr36 is known as the best glass former in this class of MGs and is commonly used as the archetype model in MD simulations (Wang et al., 2020). While other MGs are not included in this study, some common dynamic behaviors (e.g., relaxation, dynamical heterogeneity, shear band formation) are believed to be controlled by energy barriers, with structure-property relationships transferable between different MGs.

A.7. Abalation on Dataset Splits

To check our method's robustness under different dataset configurations, we performed an ablation study with two more random dataset splits. These splits use different sets of randomly sampled training graphs, one randomly sampled graph for validation, and rest for testing. The result can be found on Table 7. We see that SymGNN consistently outperforms SchNet in all the dataset splits we considered.

	Model	Train	Test
Original	SymGNN	0.8368	**0.7859**
	SchNet	0.7858	0.7588
New Split 1	SymGNN	0.8600	**0.7613**
	SchNet	0.7778	0.7583
New Split 2	SymGNN	0.8362	**0.7645**
	SchNet	0.7070	0.6948

Table 7. Performance comparison across original and new dataset splits.

A.8. Ablation on Number of Orthogonal Transformations

In our experiments, we found that the model performance is relatively stable with respect to the number of transformations. We performed an ablation study with different number of aggregated orthogonal transformations. We presented experiments results with 2, 4, 6, and 12 transformations. Although using 12 transformations led to out-of-memory (OOM) issues, we found that both 2 and 4 transformations yielded effective results, with the performance using 2 transformations even being slightly better than our reported results. We hypothesize that as long as we are sampling various orthogonal transformations from a reasonable distribution, the framework can benefit from the symmetrization module and achieve good results. The number of transformations mostly influences the convergence time rather than the performance. For example, using 2 transformations required 3000 epochs for convergence, while 4 transformations required 2300 epochs. This observation aligns with our expectations, as fewer transformations necessitate more iterations for the model to capture the necessary information from the data. The results can be found in Table 8.

Number of Transformations	0	2	4	6	12
Train	0.8736	0.8302	0.8072	0.8368	OOM
Test	0.2669	0.7901	0.7778	0.7858	OOM

Table 8. Performance ablated on different number of orthogonal transformations

A.9. Comparison with Data Augmentation

Empirically, we find that our symmetrization module can learn a condensed subspace of the orthogonal transformations corresponding to the task, which allows more effective aggregations. In this section, we provide an analysis with the subspace learned by SymGNN. We report the mean and concentration for each distribution that controlled one Euler angle after the training. For rotations, we have

$$(\mu_i, \kappa_i) = (-0.3414, 0.2985), (0.7023, 0.9146), (-1.5622, 1.3543)$$

and

$$(\mu_i, \kappa_i) = (-0.4102, 3.0350), (-0.4946, 3.0645), (-0.7043, 0.1533)$$

for the rest. We perform an estimation of how much volume we need in order to capture 80 percent of the whole probability density. We notice that Von Mises distribution with a bigger κ is approximately a Gaussian distribution with variance $\frac{1}{\kappa}$, and a Von Mises distribution with a smaller κ is approximately a uniform distribution. Therefore, we approximate the estimation using a similar density-volume estimation with four Gaussian random variables $\mathcal{N}(-1.5622, \frac{1}{1.3543}), \mathcal{N}(-0.4102, \frac{1}{3}), \mathcal{N}(-0.4946, \frac{1}{3}), \mathcal{N}(0.7023, \frac{1}{0.9146})$, and two uniform distributions between $[-\pi, \pi]$.

We found that for these four Gaussian distributions, intervals of length $2.21, 3.37, 3.37, 4.01$ around the mean can approximately yield 94.5 percent of density. Therefore, in total it will yield $0.945^4 \approx 0.8$ of density. By picking all the uniform distribution, we know that at least 80 percent of the density can be included by less than 7 percent of the density. This analysis implies the effectiveness of the learning. We should note that this estimation is rough and an overestimation, since in reality the distributions that are approximated as uniform are also more centered.

A.10. Explanation on SchNet

We compare explanations generated with different models to demonstrate the effectiveness of both our model and our explanations. We ran GNNExplainer on our strongest baseline, SchNet, and found both similarities and differences in the explanation outputs compared to ours. On one hand, the explanations are similar to the explanations generated with our model, with the most important edges being of mid-range distance to the central node (as in Figure 3). On the other hand, the comparison to TDA shows a clear difference. As we discussed in Section 6.1, we compare the edge importance identified by ML explanations to the edges captured by the TDA optimal cycles. A comparison between our explanations and SchNet explanations is shown in the table below. We observe that for SchNet explanations, there is no significant difference between the number of optimal-cycle edges participated in by those high importance edges and low importance edges (high vs. low = 1.312 vs. 1.24), whereas ours is very significant (high vs. low = 4.13 vs. 0.874). We hypothesize that this difference is due to the lack of expressivity in SchNet, as it utilizes only edge distance information and not any higher-order information, such as angles. The result can be found in Table 9.

Ave Number of Cycles	High	Medium	Low	Random
SymGNN	4.130	1.202	0.874	1.148
SchNet	1.312	1.036	1.240	1.074

Table 9. Average number of cycles involved with high/medium/low impact edges in SchNet.

A.11. Explanation Visualization

In this section, we provide more visualizations of the explanation. First, we show the pairing local explanation result for the node shown in Figure 4 in Figure 6. We note that none of the edges within the top 10 closest nodes are being selected as important edges by our explanation. Visualizations of more nodes are shown in Figure 7. In the figures, the left column presents the global version plot whereas the right column presents the local version.

Figure 6. Local and global explanation visualizations node 1501.

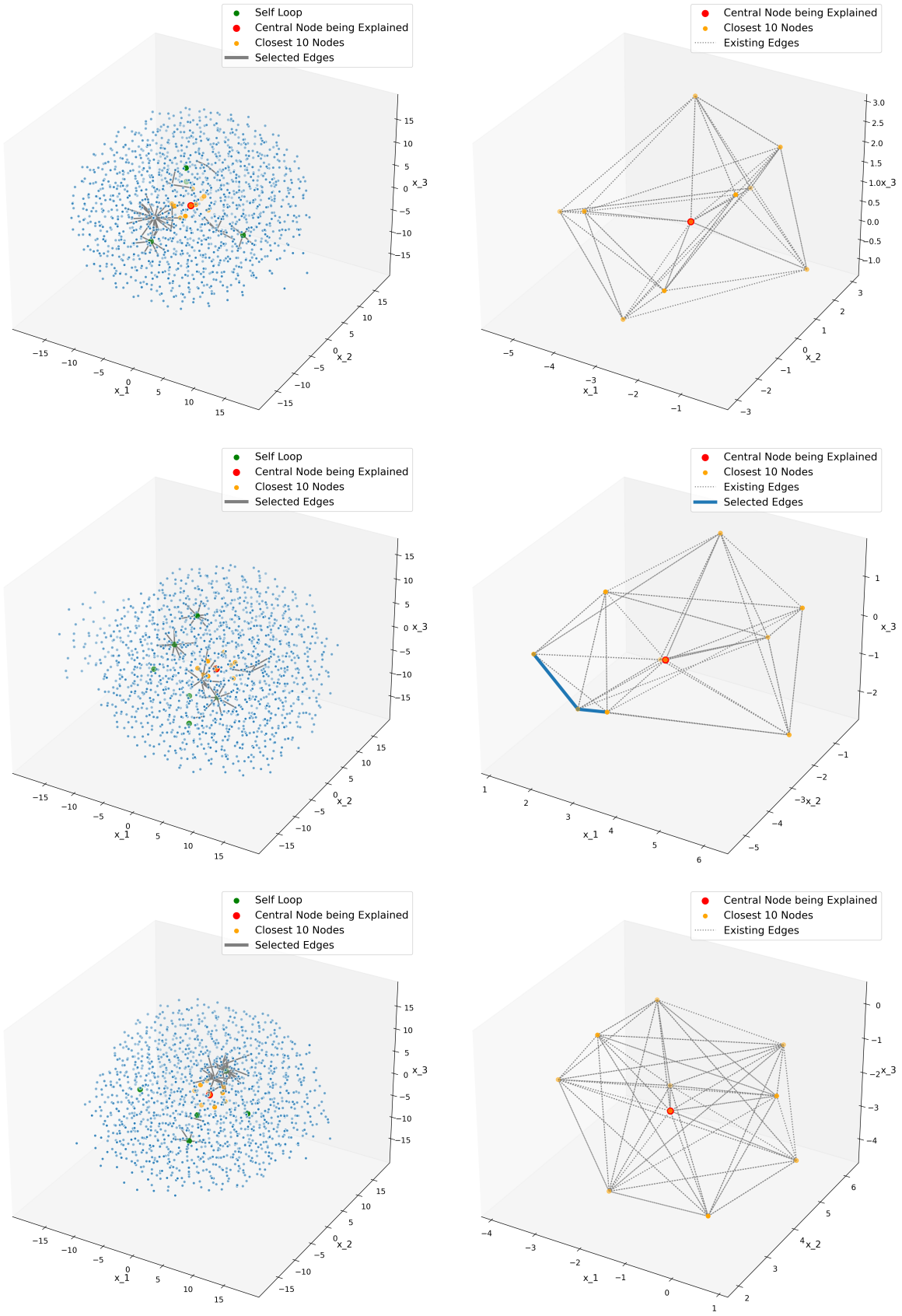

Figure 7. Local and global explanation visualizations for more nodes.