Expressive Sign Equivariant Networks for Spectral Geometric Learning

Derek Lim [1] **Joshua Robinson** [1] **Stefanie Jegelka** [1] **Haggai Maron** [2]

Abstract

Recent work has shown the utility of developing machine learning models that respect the structure and symmetries of eigenvectors. These works promote sign invariance, since for any eigenvector v the negation $-v$ is also an eigenvector. However, we show that sign invariance is theoretically limited for tasks such as building orthogonally equivariant models and learning node positional encodings for link prediction in graphs. In this work, we demonstrate the benefits of sign *equivariance* for these tasks. To obtain these benefits, we develop novel sign equivariant neural network architectures. Our models are based on a new analytic characterization of sign equivariant polynomials and thus inherit provable expressiveness properties. Controlled synthetic experiments show that our networks can achieve the theoretically predicted benefits of sign equivariant models.

1. Introduction

The need to process eigenvectors is ubiquitous in machine learning and the computational sciences. For instance, there is often a need to process eigenvectors of operators associated with manifolds or graphs, principal components (PCA) of arbitrary datasets, and eigenvectors arising from implicit or explicit matrix factorization methods. However, eigenvectors are not merely unstructured data—they have rich structure in the form of symmetries (Ovsjanikov et al., 2008).

Specifically, eigenvectors have sign and basis symmetries. An eigenvector v is sign symmetric in the sense that the sign-flipped vector $-v$ is also an eigenvector of the same eigenvalue. Basis symmetries occur when there is a repeated eigenvalue, as then there are infinitely many choices of eigenvector basis for the same eigenspace. Prior work has developed neural networks that are invariant to these symmetries (Lim et al., 2023).

The goal of this paper is to demonstrate why sign *equivariance* can be useful and to characterize fundamental expressive sign equivariant architectures. Our first contribution is to show that sign equivariant models are a natural choice for several applications, whereas sign *invariant* architectures are provably insufficient for these applications. First, we show that sign and basis invariant networks are theoretically limited in expressive power for learning multi-node representations in graphs because they learn structural node embeddings that are known to be limited for multi-node tasks (Srinivasan & Ribeiro, 2019; Zhang et al., 2021). In contrast, we show that sign equivariant models can bypass this limitation by maintaining positional information in node embeddings. Furthermore, we show that sign equivariance combined with PCA can be used to parameterize expressive orthogonally equivariant point cloud models, thus giving an efficient alternative to PCA-based frame averaging (Puny et al., 2022; Atzmon et al., 2022).

The second contribution of this work is to develop the first sign equivariant neural network architectures, with provable expressiveness guarantees. To develop our models, we derive a complete characterization of the sign equivariant polynomial functions. The form of these equivariant polynomials directly inspires our equivariant neural network architectures. Further, our architectures inherit the theoretical expressive power guarantees of the equivariant polynomials. Synthetic experiments in link prediction and n-body problems support our theory and demonstrate the utility of sign equivariant models.

1.1. Background

Let $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ be a function that takes eigenvectors $v_1, \ldots, v_k \in \mathbb{R}^n$ of an underlying matrix as input, and outputs representations $f(v_1, \ldots, v_k)$. We often concatenate the eigenvectors into a matrix $V = \begin{bmatrix} v_1, \ldots, v_k \end{bmatrix} \in \mathbb{R}^{n\times k}$, and write $f(V)$ as the application of f. For simplicity, in this work we assume the eigenvectors come from a symmetric matrix, so they are taken to be orthonormal.

Sign and basis symmetries. Eigenvectors have symmetries, because there are many possible choices of eigenvectors of a matrix. For instance, if v is a unit-norm eigenvector of a matrix, then so is the sign-flipped $-v$. If $v_1, \ldots v_m$ are an

[1]MIT CSAIL [2]NVIDIA Research. Correspondence to: Derek Lim <dereklim@mit.edu>.

Presented at the 2^{nd} Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML) at the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. 2023.

orthonormal basis of eigenvectors for the same eigenspace (meaning they all have the same eigenvalue), then there are infinitely many other choices of orthonormal basis for this eigenspace; these other choices of basis can be written as VQ, where $V = [v_1 \dots v_m] \in \mathbb{R}^{n\times m}$ and $Q \in O(m)$ is an arbitrary orthogonal matrix (Lim et al., 2023).

Sign equivariance means that if we flip the sign of an eigenvector, then the corresponding column of the output of a function f has its sign flipped. In other words, letting $\mathrm{diag}(\{-1,1\}^k)$ represent all $k\times k$ diagonal matrices with -1 or 1 on the diagonal, f is sign equivariant if

$$f(VS) = f(V)S \qquad \text{for all } S \in \mathrm{diag}(\{-1,1\}^k). \quad (1)$$

Permutation equivariance is often also a desirable property of our functions f. We say that f is *permutation equivariant* if $f(PV) = Pf(V)$ for all $n \times n$ permutation matrices P. For instance, eigenvectors of matrices associated with simple graphs of size n have such permutation symmetries, as the ordering of nodes is arbitrary.

2. Applications of Sign Equivariance

2.1. Multi-Node Representations and Link Prediction

In link prediction and multi-node prediction tasks, we typically want to learn *structural node-pair representations*, meaning adjacency-permutation equivariant functions that give a representation for each pair of nodes (Srinivasan & Ribeiro, 2019) (see Appendix B.2 for more discussion). One method to do this is to use a graph model such as a standard GNN to learn node representations z_i, and then obtain a node-pair representation for (i,j) as some function $f_{\text{decode}}(z_i, z_j)$ of z_i and z_j. However, this approach is limited because standard GNNs learn *structural node encodings* that assign the same representation to automorphic nodes (Srinivasan & Ribeiro, 2019; Zhang et al., 2021). This can be problematic since automorphic nodes can be far apart in the graph, so we may want to assign them to the different clusters (Srinivasan & Ribeiro, 2019).

One way to surpass the limitations of structural node encodings is to use *positional node embeddings*, which can assign different values to automorphic nodes. Intuitively, positional encodings capture information such as distances between nodes and global position of nodes in the graph (see (Srinivasan & Ribeiro, 2019) for a formal definition). Laplacian eigenvectors are an important example of node positional embeddings that capture much useful information of graphs (Chung, 1997).

Pitfalls of sign and basis invariance. When processing eigenvectors of matrices associated with graphs, invariance to the symmetries of the eigenvectors has been found useful (Dwivedi et al., 2022; Lim et al., 2023), especially for graph classification tasks. However, we show that exact invariance to these symmetries *removes positional information*, and thus the outputs of sign invariant or basis invariant networks are in fact *structural node encodings* (see Appendix B.2).[1] Hence, eigenvector-symmetry-invariant networks cannot learn node representations that distinguish automorphic nodes:

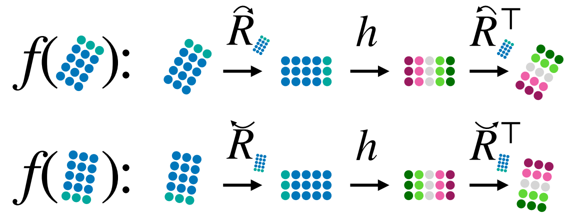

Figure 1: Using sign equivariant functions h to parameterize orthogonally equivariant $f(X) = h(XR_X)R_X^\top$. We first transform X via R_X into an orientation that is unique up to sign flips, then process XR_X using the sign equivariant model h, and finally reintegrate orientation information back into the output via $R_X^\top$.

Proposition 2.1. *Let $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n \times d_{\text{out}}}$ be a permutation equivariant function, and let $V = [v_1, \dots, v_k] \in \mathbb{R}^{n\times k}$ be k orthonormal eigenvectors of an adjacency matrix A. Let nodes i and j be automorphic, and let z_i and $z_j \in \mathbb{R}^{d_{\text{out}}}$ be their embeddings, i.e, the ith and jth row of $Z = f(V)$.*

- *If f is sign invariant and the eigenvalues associated with the v_l are distinct, then $z_i = z_j$.*
- *If f is basis invariant and $v_1, \dots, v_k$ are a basis for some number of eigenspaces of A then $z_i = z_j$.*

A novel link prediction approach via sign equivariance. The problem $z_i = z_j$ arises from the sign/basis invariances, which remove crucial positional information. We instead propose using sign *equivariant* networks (as in Section 3) to learn node representations $z_i = f(V)_{i,:} \in \mathbb{R}^k$. These representations z_i maintain positional information for each node thanks to preserving sign information. Then we use a sign invariant decoder $f_{\text{decode}}(z_i, z_j) = f_{\text{decode}}(Sz_i, Sz_j)$ for $S \in \mathrm{diag}(\{-1,1\}^k)$ to obtain node-pair representations. For instance, the commonly used $f_{\text{decode}} = \mathrm{MLP}(z_i \odot z_j)$, where $\odot$ is the elementwise product, is sign invariant. When the eigenvalues are distinct, this approach has the desired invariances (yielding structural node-pair representations) and also maintains positional information in the node embeddings; see Appendix B.2 for a proof of the invariances.

2.2. Orthogonal Equivariance

For various applications in modelling physical systems, we desire equivariance to rigid transformations. We say that a function $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ is orthogonally equivariant if $f(XQ) = f(X)Q$ for any $Q \in O(k)$, where $O(k)$ is the set of orthogonal matrices in $\mathbb{R}^{k\times k}$. Orthogonal equivariance imposes infinitely many constraints on the function f. Several works have approached this problem by reducing to a finite set of constraints using so-called Principal Component Analysis (PCA) based frames (Puny et al., 2022; Atzmon et al., 2022).

PCA-frame methods take an input $X \in \mathbb{R}^{n\times k}$, compute orthonormal eigenvectors $R_X \in O(k)$ of the covariance matrix $\operatorname{cov}(X) = (X - \frac{1}{n}\mathbf{1}\mathbf{1}^\top X)^\top (X - \frac{1}{n}\mathbf{1}\mathbf{1}^\top X)$ (assumed to have distinct eigenvalues), then average outputs of a base model h for each of the 2^k sign-flipped inputs $XR_X S$, where $S \in \operatorname{diag}(\{-1, 1\}^k)$. We instead suggest using a sign equivariant network to efficiently parameterize an efficient $O(k)$ equivariant model. For a sign equivariant network h, we define our model f to be

$$f(X) = h(XR_X)R_X^\top. \tag{2}$$

See Figure 1 for an illustration. Our approach only requires one forward pass through h, whereas frame averaging requires 2^k forward passes through a base model. The following proposition shows that f is $O(k)$ equivariant, and inherits universality properties of h.

Proposition 2.2. *Consider a domain $\mathcal{X} \subseteq \mathbb{R}^{n\times k}$ such that each $X \in \mathcal{X}$ has distinct covariance eigenvalues, and let R_X be a choice of orthonormal eigenvectors of $\operatorname{cov}(X)$ for each $X \in \mathcal{X}$. If $h : \mathcal{X} \subseteq \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ is sign equivariant, and if $f(X) = h(XR_X)R_X^\top$, then f is well defined and orthogonally equivariant.*

Moreover, if h is from a universal class of sign equivariant functions, then the f of the above form universally approximate $O(k)$ equivariant functions on $\mathcal{X}$.

3. Sign Equivariant Polynomials and Networks

3.1. Sign Equivariant Polynomials

Consider polynomials $p : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ that are sign equivariant, meaning $p(VS) = p(V)S$ for $S \in \operatorname{diag}(\{-1, 1\}^k)$. We can show (in Appendix C) that a polynomial p is sign equivariant if and only if it can be written as the elementwise product of a simple (linear) sign equivariant polynomial and a general sign invariant polynomial, followed by another linear sign equivariant map.

Theorem 3.1. *A polynomial $p : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ is sign equivariant if and only if it can be written*

$$p(V) = W^{(2)}\left((W^{(1)}V) \odot p_{\text{inv}}(V)\right) \tag{3}$$

for sign equivariant linear $W^{(2)}$ and $W^{(1)}$, and a sign invariant polynomial $p_{\text{inv}} : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$.

3.2. Sign Equivariance without Permutation Symmetries

Using Theorem 3.1, we can now develop sign equivariant architectures. We parameterize sign equivariant functions $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ as a composition of layers f_l, each of the form

$$f_l(V) = [W_1^{(l)}v_1, \ldots, W_k^{(l)}v_k] \odot \operatorname{SignNet}_l(V), \tag{4}$$

in which the $W_i^{(l)} : \mathbb{R}^n \to \mathbb{R}^{n'}$ are arbitrary linear maps, and $\operatorname{SignNet}_l : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ is sign invariant (Lim et al., 2023). In the case of $n = n' = 1$, there is a simple universal form: we can write a sign equivariant function $f : \mathbb{R}^k \to \mathbb{R}^k$ as $f(v) = v \odot \operatorname{MLP}(|v|)$, where $|v|$ is the elementwise absolute value. These two architectures are universal because they can approximate sign equivariant polynomials.

Proposition 3.2. *Functions of the form $v \mapsto v \odot \operatorname{MLP}(|v|)$ universally approximate continuous sign equivariant functions $f : \mathbb{R}^k \to \mathbb{R}^k$.*

Compositions $f_2 \circ f_1$ of functions f_l as in equation 4 universally approximate continuous sign equivariant functions $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$.

3.3. Sign Equivariance and Permutation Equivariance

To add permutation equivariance to our neural network architecture from Section 3.2, we use it within the framework of DeepSets for Symmetric Elements (DSS) (Maron et al., 2020). For a hidden dimension size of d_f, each layer $f_l : \mathbb{R}^{n\times k\times d_f} \to \mathbb{R}^{n\times k\times d_f}$ of our DSS-based sign equivariant network takes the following form on row i:

$$f_l(V)_{i,:} = f_l^{(1)}(V_{i,:}) + f_l^{(2)}\left(\sum\nolimits_{j\neq i} V_{j,:}\right), \tag{5}$$

where $f_l^{(1)}$ and $f_l^{(2)}$ are sign equivariant functions as in Section 3.2. Sometimes we take $d_f = 1$, in which case we can use the simpler $\mathbb{R}^k \to \mathbb{R}^k$ sign equivariant networks ($v \odot \operatorname{MLP}(|v|)$) as $f_l^{(1)}$ and $f_l^{(2)}$. If we have graph information, then we can do message-passing by changing the sum over $j \neq i$ to a sum over a neighborhood of node i. DSS has universal approximation guarantees (Maron et al., 2020), but they only apply for groups that act as permutation matrices, whereas the sign group $\{-1, 1\}^k$ does not. Hence, the universal approximation properties of our proposed DSS-based architecture are still an open question.

[1]When there are repeated eigenvalues, sign invariant embeddings maintain some positional information.

Table 1: Link prediction test AUC (Section 4.1).

Model	Erdős-Rényi	Barabási-Albert
GCN (constant input)	.497±.06	.705±.01
SignNet	.498±.00	.707±.00
$V_{i,:}^{\top} V_{j,:}$	.570±.01	.597±.01
$\mathrm{MLP}(V_{i,:} \odot V_{j,:})$	.614±.02	.651±.03
Sign Equivariant	**.751**±.00	**.773**±.01

4. Experiments

4.1. Link Prediction in Nearly Symmetric Graphs.

We test our models in a synthetic link prediction task. First, we either generate an Erdős-Rényi or Barabási-Albert random graph H of 1000 nodes. Then we form a larger graph G containing two disjoint copies of H, along with 1000 uniformly-randomly added edges. Without the random edges, each node in one copy of H is automorphic to a node in the other copy, so we expect many nodes to be nearly automorphic with the randomly added edges.

In Table 1, we show the link prediction performance of several models that learn structural edge representations. The methods that use eigenvectors have a sign invariant final prediction for each edge. GCN (Kipf & Welling, 2017) where the node features are all ones and SignNet (Lim et al., 2023) both completely fail on the Erdős-Rényi task (these two models map automorphic nodes to the same embedding), while our sign equivariant model outperforms all methods. We also try two eigenvector baselines that maintain node positional information, but do not update eigenvector representations: taking the dot product $V_{i,:}^{\top} V_{j,:}$ to be the logit of a link existing, or learning a simple decoder $\mathrm{MLP}(V_{i,:} \odot V_{j,:})$. Both perform substantially worse than our sign equivariant model, which shows that updating eigenvector representations is important here. See Appendix E.2 for more details.

4.2. Orthogonal Equivariance in n-body Problems

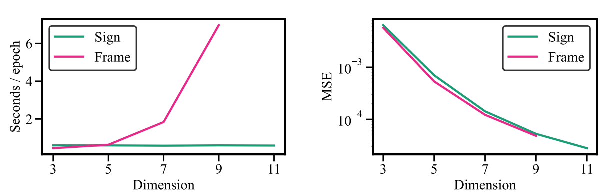

Figure 2: Sign equivariant model versus frame averaging model for n-body experiments in varying dimensions, in terms of runtime (left) and MSE (right)

In this section, we test the ability of our sign equivariant models to parameterize orthogonally equivariant functions on point clouds, as outlined in Section 2.2. For this purpose, we consider simulating n-body problems, following the setup in Fuchs et al. (2020) and building on the code from Puny et al. (2022), except that we generalize their three-dimensional experiments to general $d \geq 3$.

Figure 2 illustrates the runtime and MSE. The sign equivariant model scales well with dimension—the time-per-epoch is nearly constant as we increase the dimension. In contrast, frame averaging suffers from the expected exponential slowdown with dimension, and runs out of memory on a 32GB V100 GPU for $d = 11$. Considering the MSE, the equivariant model's performance closely follows that of frame averaging: the sign equivariant model has an MSE of .00646 (compared to .00575 of frame averaging), which means that the sign equivariant model outperforms all of the other methods tested in Puny et al. (2022).

Acknowledgements

We would like to thank Yaron Lipman for help on an earlier version of this project. We would like to thank Maks Ovsjanikov for noting that basis invariant models give automorphic nodes the same representation, and also Johannes Lutzeyer and Michael Murphy for helpful comments. DL is funded by an NSF Graduate Fellowship. SJ acknowledges support from NSF award CCF-2112665 and NSF award 2134795.

References

Atzmon, M., Nagano, K., Fidler, S., Khamis, S., and Lipman, Y. Frame averaging for equivariant shape space learning. In *CVPR*, 2022.

Chung, F. *Spectral graph theory*. American Mathematical Society, 1997.

Dwivedi, V. P., Joshi, C. K., Laurent, T., Bengio, Y., and Bresson, X. Benchmarking graph neural networks. In *JMLR*, 2022.

Fuchs, F., Worrall, D., Fischer, V., and Welling, M. SE(3)-transformers: 3D roto-translation equivariant attention networks. In *NeurIPS*, 2020.

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

Lim, D., Robinson, J. D., Zhao, L., Smidt, T., Sra, S., Maron, H., and Jegelka, S. Sign and basis invariant networks for spectral graph representation learning. In *ICLR*, 2023.

Maron, H., Litany, O., Chechik, G., and Fetaya, E. On learning sets of symmetric elements. In *ICML*, pp. 6734–6744. PMLR, 2020.

Ovsjanikov, M., Sun, J., and Guibas, L. Global intrinsic symmetries of shapes. In *Computer graphics forum*, volume 27, pp. 1341–1348. Wiley Online Library, 2008.

Puny, O., Atzmon, M., Ben-Hamu, H., Smith, E. J., Misra, I., Grover, A., and Lipman, Y. Frame averaging for invariant and equivariant network design. In *ICLR*, volume 10, 2022.

Srinivasan, B. and Ribeiro, B. On the equivalence between positional node embeddings and structural graph representations. In *ICLR*, 2019.

Zhang, M., Li, P., Xia, Y., Wang, K., and Jin, L. Labeling trick: A theory of using graph neural networks for multi-node representation learning. In *NeurIPS*, 2021.

A. Related Work

Structural and Positional Representations. Especially for link prediction, the need for structural node-pair representations that are not obtained from structural node representations has been discussed in several works (Srinivasan & Ribeiro, 2019; Zhang et al., 2021; Cotta et al., 2023). As such, various methods have been developed for learning structural node-pair representations that incorporate node positional information. SEAL and other labeling-trick based methods (Zhang & Chen, 2018; Zhang et al., 2021) use added node features depending on the node-pair that we want a representation of. This is empirically successful in many tasks, but typically requires a separate subgraph extraction and forward pass through a GNN for each node-pair under consideration. Distance encoding (Li et al., 2020) uses relative distances between nodes to capture positional information. PEG (Wang et al., 2022) similarly maintains positional information by using eigenvector distances between nodes in each layer of a GNN, but does not update eigenvector representations. Identity-aware GNNs (You et al., 2021) and Neural Bellman-Ford Networks (Zhu et al., 2021) learn pair representations by conditioning on a source node from the pair.

Eigenvectors as Graph Positional Encodings. When using eigenvectors of graphs as node positional encodings for graph models like GNNs and Graph Transformers, many works have noted the need to address the sign ambiguity of the eigenvectors. This is often done by encouraging sign invariance through data augmentation—the signs of the eigenvectors are chosen randomly in each iteration of training (Dwivedi et al., 2022; Kreuzer et al., 2021; Mialon et al., 2021; Kim et al., 2022; He et al., 2022; Müller et al., 2023). In contrast, SignNet (Lim et al., 2023) enforces exact sign invariance, by processing eigenvectors with a sign invariant neural architecture; this approach has been taken by some recent works (Rampasek et al., 2022; Geisler et al., 2023; Murphy et al., 2023).

Equivariant Neural Network Design. Equivariant neural network architectures have been proposed for various types of data and symmetry groups. A common paradigm is to interleave equivariant linear maps and equivariant pointwise nonlinearities (Wood & Shawe-Taylor, 1996; Cohen & Welling, 2016; 2017; Ravanbakhsh et al., 2017; Maron et al., 2018; Kondor & Trivedi, 2018; Finzi et al., 2021; Bronstein et al., 2021); this is often used when the group acts as some subset of the permutation matrices. However, the sign group does not act as permutation matrices, and as we explained above this approach is not expressive for sign equivariant models. More similarly to our approach, many equivariant machine learning works heavily leverage invariant or equivariant polynomials (or other equivariant nonlinear functions). These works include polynomials as operations within a network (Thomas et al., 2018; Puny et al., 2023), add polynomials as features (Yarotsky, 2022; Villar et al., 2021), build networks that take a similar form to equivariant polynomials (Villar et al., 2021), and/or analyze neural network expressive power by determining which equivariant polynomials a given architecture can compute (Zaheer et al., 2017; Segol & Lipman, 2019; Maron et al., 2019; Puny et al., 2023).

B. Applications of Sign Equivariance

B.1. Edge Representations and Link Prediction

B.1.1. Proof of Proposition 2.1

Proposition 2.1. Let $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n \times d_{\text{out}}}$ be a permutation equivariant function, and let $V = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$ be k orthonormal eigenvectors of an adjacency matrix A. Let nodes i and j be automorphic, and let z_i and $z_j \in \mathbb{R}^{d_{\text{out}}}$ be their embeddings, i.e, the ith and jth row of $Z = f(V)$.

- If f is sign invariant and the eigenvalues associated with the v_l are distinct, then $z_i = z_j$.
- If f is basis invariant and $v_1, \ldots, v_k$ are a basis for some number of eigenspaces of A then $z_i = z_j$.

Proof. We only prove the basis invariance claim, as the sign invariance claim is a special case; basis invariance is sign invariance when eigenvalues are distinct.

Let $P \in \mathbb{R}^{n\times n}$ be a permutation matrix associated to an automorphism that maps node i to node j, so $PAP^\top = A$ and $Pe_i = e_j$, where e_l is the lth standard basis vector. Let $V_t = [v_{r_1}, \ldots, v_{r_{d_t}}]$ be the matrix whose columns are the eigenvectors v_{r_l} that are associated to eigenvalue λ_i. The columns of V_t are thus an orthonormal basis for the eigenspace associated to λ_t. Note that for any of these eigenvectors, we have

$$A(Pv_{r_l}) = PAP^\top(Pv_{r_l}) = PAv_{r_l} = P\lambda_i v_{r_l} = \lambda_t(Pv_{r_l}), \tag{6}$$

so Pv_{r_l} is also an eigenvector of A with eigenvalue λ_t. As P is orthogonal, note that $Pv_{r_1}, \dots, Pv_{r_{d_t}}$ is still an orthonormal basis of the eigenspace. Thus, there exists an orthogonal matrix $Q_t \in \mathbb{R}^{d_t \times d_t}$ such that $PV_t = V_t Q_t$—see Lim et al. (2023).

Repeat the above argument to get such a Q_t for each of the eigenbases $V_1, \dots, V_l$. We can then see that

$$\begin{aligned}
z_j &= f(V_1, \dots, V_l)_{j,:} \\
&= f(V_1 Q_1, \dots, V_l Q_l)_{j,:} && \text{basis invariance} \\
&= f(PV_1, \dots, PV_l)_{j,:} && \text{choice of } Q_t \\
&= (Pf(V_1, \dots, V_l))_{j,:} && \text{permutation equivariance} \\
&= f(V_1, \dots, V_l)_{i,:} && \text{choice of } P \\
&= z_i.
\end{aligned}$$

So we are done. □

B.2. Sign Invariance and Structural Node or Node-Pair Encodings

In this section, we show that when the eigenvalues $\lambda_1, \dots, \lambda_k$ are distinct, then sign invariant functions of the orthonormal eigenvectors $v_1, \dots, v_k$ give structural node or node-pair representations. This can also be generalized in a straightforward way to larger tuples of nodes beyond pairs, though we only consider nodes and node-pairs for ease of exposition. First, we give formal definitions.

Definition B.1 (Structural Representations (Srinivasan & Ribeiro, 2019))**.** Let $A \in \mathbb{R}^{n \times n}$ be the adjacency matrix of a graph on node set $\{1, \dots, n\}$.

A function $f : \mathbb{R}^{n \times n} \to \mathbb{R}^n$ is a node structural representation if $f(PAP^\top) = Pf(A)$ for all $n \times n$ permutation matrices P.

A function $f : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ is a node-pair structural representation if $f(PAP^\top) = Pf(A)P^\top$ for all $n \times n$ permutation matrices P.

Importantly, these structural representations are permutation equivariant functions of adjacency matrices, not arbitrary matrices. For each adjacency matrix A, let $V(A) = [v_1(A), \dots, v_k(A)]$ be a choice of orthonormal eigenvectors for the first k eigenvalues $\lambda_1(A), \dots, \lambda_k(A)$. We assume in this section that these first k eigenvalues are distinct for all A under consideration, so $V(A)$ is defined up to sign flips. Let $h : \mathbb{R}^{n \times k} \to \mathbb{R}^n$ be a permutation equivariant function of sets, so $h(PX) = Ph(X)$ for all permutations matrices P. Then of course $h(PV(A)) = Ph(V(A))$, but this does not make h a node structural encoding. This is because $A \mapsto h(V(A))$ is in general not a well-defined function of the adjacency, since the choice of $V(A)$ is not well-defined (the choices of sign are arbitrary). If we constrain h to not depend on the signs (sign invariance), or to depend on the signs in a predictable way (sign equivariance), then we can compute structural node or node-pair encodings from eigenvectors.

We capture these observations in the below proposition. First, we define three types of functions:

- Let $f_{\text{node}} : \mathbb{R}^{n \times k} \to \mathbb{R}^n$ be sign invariant and permutation equivariant; that is, $f_{\text{node}}(Pv_1 s_1, \dots, Pv_k s_k) = Pf_{\text{node}}(v_1, \dots, v_k)$ for $s_i \in \{-1, 1\}$ and P a permutation matrix.
- Let $f_{\text{decode}} : \mathbb{R}^{2 \times k} \to \mathbb{R}$ be sign invariant; that is, $f_{\text{decode}}(Sz_i, Sz_j) = f_{\text{decode}}(z_i, z_j)$ for $S \in \operatorname{diag}(\{-1, 1\}^k)$.
- Let $f_{\text{equiv}} : \mathbb{R}^{n \times k} \to \mathbb{R}^{n \times k}$ be a permutation equivariant and sign equivariant function; that is, $f_{\text{equiv}}(PV(A)S) = Pf_{\text{equiv}}(V(A))S$ for $S \in \operatorname{diag}(\{-1, 1\}^k)$ and P a permutation matrix.

Proposition B.2. *Let $\mathcal{A} \subseteq \mathbb{R}^{n \times n}$ denote the matrices with distinct first-k eigenvalues. For $A \in \mathcal{A}$, let $V(A) = [v_1(A), \dots, v_k(A)]$ be a choice of orthonormal eigenvectors of A, associated to the first-k (distinct) eigenvalues $\lambda_1(A), \dots, \lambda_k(A)$. Then*

(a) The map $q_{\text{node}} : \mathcal{A} \to \mathbb{R}^n$ given by $q_{\text{node}}(A)_i = f_{\text{node}}\left(f_{\text{equiv}}(V(A))\right)_i$ is well-defined and gives a structural node representation.

(b) The map $q_{\text{pair}} : \mathcal{A} \to \mathbb{R}^{n \times n}$ defined by $q_{\text{pair}}(A)_{i,j} = f_{\text{decode}}\left(f_{\text{equiv}}(V(A))_{i,:}, f_{\text{equiv}}(V(A))_{j,:}\right)$ is well-defined and gives a structural node-pair representation.

Note that the identity mapping $V(A) \mapsto V(A)$ is permutation equivariant and sign equivariant, so using f_{node} or f_{decode} directly on eigenvectors also gives structural representations. The statement (b) means that our link prediction pipeline with sign equivariant node features and sign invariant decoding produces structural node-pair representations.

Proof. **Part (a)** We first show that $q_{\text{node}} : \mathcal{A} \to \mathbb{R}^n$ is well-defined. Suppose we had another choice of eigenvectors, so the eigenvectors we input are $V(A)S$ for some $S \in \text{diag}(\{-1, 1\}^k)$. Then

$$f_{\text{node}}\left(f_{\text{equiv}}(V(A)S)\right) = f_{\text{node}}\left(f_{\text{equiv}}(V(A))S\right) = f_{\text{node}}\left(f_{\text{equiv}}(V(A)\right), \tag{7}$$

where the first equality is by sign equivariance, and the second equality by sign invariance. Thus, the value of $q_{\text{node}}(A)$ is unchanged.

Now, let P be any permutation matrix. Then for each eigenvector $v_i(A)$, $i \in [k]$, we have $(PAP^\top)Pv_i(A) = PAv_i(A) = \lambda_i(A)Pv_i(A)$, so $Pv_i(A)$ is an eigenvector of $PAP^\top$ associated to $\lambda_i(A) = \lambda_i(PAP^\top)$. Hence, we denote $v_i(PAP^\top) = Pv_i(A)$ (the choice of sign does not matter as q does not depend on the sign. Now, we have that

$$\begin{aligned} q_{\text{node}}(PAP^\top) &= f_{\text{node}}\left(f_{\text{equiv}}(V(PAP^\top))\right) && (8)\\ &= f_{\text{node}}\left(f_{\text{equiv}}(PV(A))\right) && (9)\\ &= Pf_{\text{node}}\left(f_{\text{equiv}}(V(A))\right) && (10)\\ &= Pq_{\text{node}}(A) && (11)\end{aligned}$$

where the second to last equality is by permutation equivariance of f_{node} and f_{equiv}.

Part (b) That $q_{\text{pair}} : \mathcal{A} \to \mathbb{R}^{n \times n}$ is well-defined follows from a similar argument to the q_{node} case. Let P be a permutation matrix, and $\sigma : [n] \to [n]$ its underlying permutation. We compute that

$$\begin{aligned} q_{\text{pair}}(PAP^\top)_{i,j} &= f_{\text{decode}}\left(f_{\text{equiv}}(V(PAP^\top))_{i,:}, f_{\text{equiv}}(V(PAP^\top))_{j,:}\right) && (12)\\ &= f_{\text{decode}}\left(f_{\text{equiv}}(PV(A))_{i,:}, f_{\text{equiv}}(PV(A))_{j,:}\right) && (13)\\ &= f_{\text{decode}}\left([Pf_{\text{equiv}}(V(A))]_{i,:}, [Pf_{\text{equiv}}(V(A))]_{j,:}\right) && (14)\\ &= f_{\text{decode}}\left(f_{\text{equiv}}(V(A))_{\sigma^{-1}(i),:}, f_{\text{equiv}}(V(A))_{\sigma^{-1}(j),:}\right) && (15)\\ &= q_{\text{pair}}(A)_{\sigma^{-1}(i),\sigma^{-1}(j)} && (16)\\ &= (Pq_{\text{pair}}(A)P^\top)_{i,j} && (17)\end{aligned}$$

□

B.2.1. Sign Equivariance is Provably More Expressive for Link Prediction

Our arguments in Section 2.1 explain why we can expect sign equivariant models to be more powerful than sign invariant models in link prediction. To give a theoretically rigorous explanation, here we provide an example where sign equivariant models can provably compute more expressive link representations than sign invariant models.

Consider a cycle graph C_{2k} for some even length $2k$, where $k \geq 3$. All nodes are automorphic in this graph, so any model based on structural node representations must assign the same representation to each node-pair. For instance, consider the eigenvalue -2 of the adjacency matrix, which is a simple eigenvalue with eigenvector $[1, -1, 1, -1, \ldots, 1, -1]$ (Lee et al., 1992). Then a sign invariant model will lose the sign information and map each node to the same encoding, which means that each node-pair will also have the same encoding. However, a sign equivariant model can preserve the sign of each node (for instance by learning the identity function). Then for any pair of nodes that are one hop away, it can take a dot product to compute the pair representation -1, whereas it can take a dot product between any nodes that are two hops away to compute the pair representation 1. Of course, using more eigenvectors would allow for more complex representations to be computed.

B.3. Proof of Proposition 2.2, Orthogonal Equivariance via Sign Equivariance

Proposition 2.2. Consider a domain $\mathcal{X} \subseteq \mathbb{R}^{n \times d}$ such that each $X \in \mathcal{X}$ has distinct covariance eigenvalues, and let R_X be a choice of orthonormal eigenvectors of $\text{cov}(X)$ for each $X \in \mathcal{X}$. If $h : \mathcal{X} \subseteq \mathbb{R}^{n \times d} \to \mathbb{R}^{n \times d}$ is sign equivariant, and if $f(X) = h(XR_X)R_X^\top$, then f is well defined and orthogonally equivariant.

Moreover, is h is from a universal class of sign equivariant functions, then the f of the above form universally approximate $O(k)$ equivariant functions on $\mathcal{X}$.

Proof. First, we show that f is well defined. R_X is only unique up to sign flips, as $R_X S$ is an orthonormal set of eigenvectors of $\text{cov}(X)$ for $S \in \text{diag}(\{-1,1\}^k)$. However, no matter the choice of signs, $f(X)$ takes the same value, since

$$\begin{aligned} h(XR_XS)(R_XS)^\top &= h(XR_XS)S^\top R_X^\top && (18)\\ &= h(XR_X)SS^\top R_X^\top && \text{sign equivariance} \quad (19)\\ &= h(XR_X)R_X^\top. && (20) \end{aligned}$$

Next, we show that f is $O(k)$ equivariant. Let $Q \in O(k)$ be any orthogonal matrix. Note that

$$\text{cov}(XQ) = \left(XQ - \frac{1}{n}\mathbf{1}\mathbf{1}^\top XQ\right)^\top \left(XQ - \frac{1}{n}\mathbf{1}\mathbf{1}^\top XQ\right) = Q^\top \text{cov}(X) Q. \tag{21}$$

Thus, $Q^\top R_X$ is an orthonormal set of eigenvectors of $\text{cov}(XQ)$. This means that there is a choice of signs $S \in \text{diag}(\{-1,1\}^k)$ such that $Q^\top R_X S = R_{XQ}$. Hence, we have that

$$\begin{aligned} f(XQ) &= h(XQR_{XQ})R_{XQ}^\top && (22)\\ &= h(XQQ^\top R_X S)(Q^\top R_X S)^\top && (23)\\ &= h(XR_X)SS^\top R_X^\top Q && \text{sign equivariance} \quad (24)\\ &= h(XR_X)R_X^\top Q && (25)\\ &= f(X)Q^\top, && (26) \end{aligned}$$

so f is $O(k)$ equivariant.

Universal Approximation. Our proof of the universality of this class of functions builds on the proof of the universality of frame averaging (Puny et al., 2022). Let f_{target} be a continuous $O(k)$ equivariant function and let $\epsilon > 0$ be a desired approximation accuracy. Then f_{target} is also sign equivariant (as the sign matrices $S \in \text{diag}(\{-1,1\}^k)$ are orthogonal).

Hence, by sign equivariant universality, we can choose a sign equivariant h such that $\|h(X) - f_{\text{target}}(X)\| < \epsilon$ for all $X \in \mathcal{X}$ (where $\|\cdot\|$ is the Frobenius norm). Define the $O(k)$ equivariant $f(X) = h(XR_X)R_X^\top$. Then for all $X \in \mathcal{X}$ we have that

$$\begin{aligned} \|f_{\text{target}}(X) - f(X)\| &= \left\|f_{\text{target}}(X) - h(XR_X)R_X^\top\right\| && (27)\\ &= \left\|f_{\text{target}}(X)R_XR_X^\top - h(XR_X)R_X^\top\right\| && R_X \text{ orthogonal} \quad (28)\\ &= \left\|f_{\text{target}}(XR_X)R_X^\top - h(XR_X)R_X^\top\right\| && \text{orthogonal equivariance} \quad (29)\\ &= \|f_{\text{target}}(XR_X) - h(XR_X)\| && R_X \text{ orthogonal} \quad (30)\\ &< \epsilon. && (31) \end{aligned}$$

So f approximates f_{target} within ϵ accuracy on $\mathcal{X}$, and we are done. □

C. Characterization of Sign Equivariant Polynomials

In this Appendix, we characterize the form of the sign equivariant polynomials. This is useful, because for a finite group, equivariant polynomials universally approximate equivariant continuous functions (Yarotsky, 2022); thus, if a model universally approximates equivariant polynomials, then it universally approximates equivariant continuous functions. Using equivariant polynomials to analyze or develop equivariant machine learning models has been done successfully in many contexts (Zaheer et al., 2017; Yarotsky, 2022; Segol & Lipman, 2019; Dym & Maron, 2021; Maron et al., 2019; 2020; Villar et al., 2021; Dym & Gortler, 2022; Puny et al., 2023).

C.1. Sign Equivariant Linear Map Characterization

Here, we prove a result characterizing the form of the equivariant linear maps.

Lemma C.1. *A linear map $W : \mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$ is sign equivariant if and only if it can be written as*

$$W(X) = \begin{bmatrix} W_1 X_1 & \dots & W_k X_k \end{bmatrix} \tag{32}$$

for some linear maps $W_1, \dots, W_k : \mathbb{R}^n \to \mathbb{R}^{n'}$, where $X_i \in \mathbb{R}^n$ is the ith column of $X \in \mathbb{R}^{n \times k}$.

Proof. For one direction, suppose W can be written as in equation 32. To see that W is sign equivariant, note that for any $S \in \mathrm{diag}(\{-1, 1\}^k)$, we have

$$W(XS) = \begin{bmatrix} s_1 W_1 X_1 & \dots & s_k W_k X_k \end{bmatrix} = \begin{bmatrix} W_1 X_1 & \dots & W_k X_k \end{bmatrix} S = W(X)S. \tag{33}$$

For the other direction, let W be a sign equivariant linear map. For any $i' \in [n']$ and $j' \in [k]$, we can write the action of W as

$$W(X)_{i',j'} = \sum_{i=1}^{n} \sum_{j=1}^{k} W_{i',j'}^{i,j} X_{i,j}, \tag{34}$$

where $W_{i',j'}^{i,j} \in \mathbb{R}$ are coefficients representing the linear map. Let $c \neq j'$ be a column that is not j'. Further, for any row $l \in [n]$, let $\tilde{X} \in \mathbb{R}^{n \times k}$ be such that $\tilde{X}_{l,c} = 1$, and $\tilde{X}$ is zero elsewhere. Then we have that

$$W(\tilde{X})_{i',j'} = W_{i',j'}^{l,c}. \tag{35}$$

Now, let $S \in \mathrm{diag}(\{-1, 1\}^k)$ have a -1 in the j'th column and a 1 elsewhere. Then $\tilde{X}S = \tilde{X}$. This implies that

$$\begin{align} W_{i',j'}^{l,c} &= W(\tilde{X})_{i',j'} \tag{36} \\ &= W(\tilde{X}S)_{i',j'} \tag{37} \\ &= -W(\tilde{X})_{i',j'} \tag{38} \\ &= -W_{i',j'}^{l,c}, \tag{39} \end{align}$$

where in the second to last equality we used sign equivariance. This implies that $W_{i',j'}^{l,c} = 0$.

Hence, for any $i' \in [n'], j' \in [k']$, we have that $W(X)_{i',j'}$ only depends on $X_{j'}$, so we are done. □

C.2. Sign Invariant Polynomials $\mathbb{R}^k \to \mathbb{R}$

For simplicity, we start with the case of sign invariant polynomials $p : \mathbb{R}^k \to \mathbb{R}$. The sign equivariant polynomials take a very similar form. We can write any polynomial from $\mathbb{R}^k$ to $\mathbb{R}$ in the form

$$p(v) = \sum_{d_1, \dots, d_k = 0}^{D} \mathbf{W}_{d_1, \dots, d_k} v_1^{d_1} \cdots v_k^{d_k} \tag{40}$$

for some coefficients $\mathbf{W}_{d_1, \dots, d_k} \in \mathbb{R}$ and some $D \in \mathbb{N}$. Sign invariance tells us that for any $S = \mathrm{diag}(s_1, \dots, s_k) \in \mathrm{diag}(\{-1, 1\}^k)$, we must have

$$\sum_{d_1, \dots, d_k = 0}^{D} \mathbf{W}_{d_1, \dots, d_k} v_1^{d_1} \cdots v_k^{d_k} = p(v) = p(Sv) = \sum_{d_1, \dots, d_k = 0}^{D} \mathbf{W}_{d_1, \dots, d_k} s_1^{d_1} \cdots s_k^{d_k} v_1^{d_1} \cdots v_k^{d_k}. \tag{41}$$

This holds for any $v \in \mathbb{R}^k$, so for all choices of $d_1, \dots, d_k$ we must have

$$\mathbf{W}_{d_1, \dots, d_k} = s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}_{d_1, \dots, d_k}, \quad \text{for all } (s_1, \dots, s_k) \in \{-1, 1\}^k. \tag{42}$$

Note that $s_i^{d_i} = 1$ if d_i is an even number. Hence, there are no constraints on $\mathbf{W}_{d_1,\ldots,d_k}$ if all d_i are even. On the other hand, suppose d_j is odd for some j. Let $s_i = 1$ for $i \neq j$ and $s_j = -1$. Then the constraint says that $\mathbf{W}_{d_1,\ldots,d_k} = -\mathbf{W}_{d_1,\ldots,d_k}$, so we must have $\mathbf{W}_{d_1,\ldots,d_k} = 0$. To summarize, we have

$$\mathbf{W}_{d_1,\ldots,d_k} = \begin{cases} \text{free} & d_i \text{ even for each } i \\ 0 & \text{else} \end{cases} \tag{43}$$

Where being free means that the coefficient may take any value in $\mathbb{R}$. Thus, any sign invariant p only has terms where each variable v_i is raised to an even power. It is also easy to see that any polynomial p where each variable v_i is raised to only even powers is sign invariant, so we have the following proposition:

Proposition C.2. *A polynomial $p : \mathbb{R}^k \to \mathbb{R}$ is sign invariant if and only if it can be written*

$$p(v) = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}, \tag{44}$$

for some coefficients $\mathbf{W}_{d_1,\ldots,d_k} \in \mathbb{R}$ and $D \in \mathbb{N}$.

In other words, p is sign invariant if and only if there exists a polynomial $q : \mathbb{R}^k \to \mathbb{R}$ such that $p(v) = q(v_1^2, \ldots, v_k^2)$.

C.3. Sign Equivariant Polynomials $\mathbb{R}^k \to \mathbb{R}^k$

The case of sign equivariant polynomials $p : \mathbb{R}^k \to \mathbb{R}^k$ is very similar. For $l \in [k]$, the lth output dimension of any polynomial $p : \mathbb{R}^k \to \mathbb{R}^k$ can be written

$$p(v)_l = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{d_1} \cdots v_k^{d_k}, \tag{45}$$

where $\mathbf{W}^{(l)}_{d_1,\ldots,d_k} \in \mathbb{R}$ are coefficients (note the extra l index, so there are k times more coefficients than in the invariant case). By sign equivariance, we have

$$s_l \cdot p(v)_l = p(Sv)_l \tag{46}$$

$$s_l \cdot \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{d_1} \cdots v_k^{d_k} = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} s_1^{d_1} \cdots s_k^{d_k} v_1^{d_1} \cdots v_k^{d_k}. \tag{47}$$

As this holds for all inputs $v \in \mathbb{R}^k$, we have the following constraints on the coefficients:

$$s_l \mathbf{W}^{(l)}_{d_1,\ldots,d_k} = s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} \tag{48}$$

$$\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = s_l \cdot s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}^{(l)}_{d_1,\ldots,d_k}, \tag{49}$$

where we use the fact that $s_l = 1/s_l$ since $s_l \in \{-1, 1\}$. If d_j is odd for $j \neq l$, then similarly to the invariant case, we can take $s_i = 1$ for $i \neq j$ and $s_j = -1$ in the above equation to see that $\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = 0$. If d_l is even, then $d_l + 1$ is odd, so we have that $\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = 0$ by the same argument. Thus, we must have

$$\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = \begin{cases} \text{free} & d_l \text{ odd, and } d_i \text{ even for each } i \neq l \\ 0 & \text{else} \end{cases}. \tag{50}$$

Thus, the lth entry $p(v)_l$ only contains monomials of the term $v_1^{2d_1} \cdots v_l^{2d_l+1} \cdots v_k^{2d_k}$, where each term besides v_l is raised to an even power. We can factor out a v_l and write such terms as $v_l \cdot v_1^{2d_1} \cdots v_k^{2d_k}$. It is also easy to see that any polynomial with monomials only of this form is sign equivariant. Thus, we have proven Proposition C.3.

Proposition C.3. *A polynomial $p : \mathbb{R}^k \to \mathbb{R}^k$ is sign equivariant if and only if it can be written*

$$p(v)_l = v_l \cdot \left(\sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k} \right). \tag{51}$$

In vector format, p is sign equivariant if and only if it can be written as $p(v) = v \odot p_{\text{inv}}(v)$ for a sign invariant polynomial $p_{\text{inv}} : \mathbb{R}^k \to \mathbb{R}^k$.

C.4. Sign Equivariant Polynomials $\mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$

Finally, we will handle the case of polynomials $p : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ equivariant to $\mathrm{diag}(\{-1,1\}^k)$. This is the case we most often deal with in practice, when we have input $V = \begin{bmatrix} v_1 & \dots & v_k \end{bmatrix}$ for k eigenvectors $v_i \in \mathbb{R}^n$ of some $n \times n$ matrix. For $a \in [n']$ and $b \in [k]$, the (a,b)th output of a polynomial $\mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ is

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} \prod_{i=1}^{n} \prod_{j=1}^{k} V_{i,j}^{d_{i,j}}, \tag{52}$$

where the sum ranges over $d_{i,j} \in \{0,\dots,D\}$ for $i \in [n]$ and $j \in [k]$, and $\mathbf{d} = (d_{1,1},\dots,d_{n,1},d_{1,2},\dots,d_{n,k})$ is a shorthand to index coefficients $\mathbf{W}_{\mathbf{d}}^{(a,b)} \in \mathbb{R}$. By sign equivariance, we have that:

$$s_b \cdot p(V)_{a,b} = p(VS)_{a,b} \tag{53}$$

$$s_b \cdot \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} \prod_{i=1}^{n} \prod_{j=1}^{k} V_{i,j}^{d_{i,j}} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} s_1^{\tilde{d}_1} \cdots s_k^{\tilde{d}_k} \prod_{i=1}^{n} \prod_{j=1}^{k} V_{i,j}^{d_{i,j}}, \tag{54}$$

where $\tilde{d}_j = \sum_{i'=1}^{n} d_{i',j}$ is the number of times that an entry from column j of V appears in the product $\prod_{i=1}^{n} \prod_{j=1}^{k} V_{i,j}^{d_{i,j}}$. As this holds over all V, we thus have that

$$\mathbf{W}_{\mathbf{d}}^{(a,b)} = s_b \cdot s_1^{\tilde{d}_1} \cdots s_k^{\tilde{d}_k} \cdot \mathbf{W}_{\mathbf{d}}^{(a,b)}. \tag{55}$$

By analogous arguments to the previous subsections, if $\tilde{d}_j$ is odd for $j \neq b$, we have that the $\mathbf{W}_{\mathbf{d}}^{(a,b)} = 0$. Likewise, if $\tilde{d}_b$ is even, we have $\mathbf{W}_{\mathbf{d}}^{(a,b)} = 0$. Thus, the constraint on $\mathbf{W}$ is

$$\mathbf{W}_{\mathbf{d}}^{(a,b)} = \begin{cases} \text{free} & \sum_i d_{i,b} \text{ odd, and } \sum_i d_{i,j} \text{ even for each } j \neq b \\ 0 & \text{else} \end{cases}. \tag{56}$$

In particular, this means that the only nonzero terms in the sum that defines $p(V)_{a,b}$ have an even number of entries from column j for $j \neq b$, and an odd number of entries from column b. Thus, each term can be written as $V_{i_{\mathbf{d}},b} \cdot p_{\mathrm{inv}}(V)_{\mathbf{d}}$ for some index $i_{\mathbf{d}} \in [n]$ and sign invariant polynomial p_{inv}. Moreover, it can be seen that any polynomial that only has terms of this form is sign equivariant. Thus, we have shown the following proposition:

Proposition C.4. *A polynomial $p : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ is sign equivariant if and only if it can be written as*

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} V_{i_{\mathbf{d}},b} \cdot p_{\mathrm{inv}}(V)_{\mathbf{d}}, \tag{57}$$

where p_{inv} is a sign invariant polynomial, the sum ranges over all $\mathbf{d}$, and $i_{\mathbf{d}} \in [n]$ for each $\mathbf{d}$.

Now, we show that this implies Theorem 3.1. In particular, we will write p in the form

$$p(V) = W^{(2)} \left((W^{(1)} V) \odot q_{\mathrm{inv}}(V) \right), \tag{58}$$

for sign equivariant linear maps $W^{(2)}$ and $W^{(1)}$, and a sign equivariant polynomial q_{inv}. To do so, let $\tilde{D}$ denote the number of all possible $\mathbf{d}$ that the sum in equation 57 ranges over. We take $W^{(1)} : \mathbb{R}^{n\times \tilde{k}} \to \mathbb{R}^{\tilde{D}n'\times k}$ and $W^{(2)} : \mathbb{R}^{\tilde{D}n'\times k} \to \mathbb{R}^{n'\times k}$. These sign equivariant linear maps have to act independently on each column of their input, so $W^{(1)}V = [W_1^{(1)} v_1, \dots W_k^{(1)} v_k]$ for linear maps $W_i^{(1)} : \mathbb{R}^n \to \mathbb{R}^{\tilde{D}n'}$. We define $W_b^{(1)}$ to be the linear map such that $(W_b^{(1)} v_b)_{\mathbf{d},a} = \mathbf{W}_{\mathbf{d}}^{(a,b)} V_{i_{\mathbf{d}},b}$ for $a \in [n']$. For the sign invariant polynomial q_{inv}, we take $q_{\mathrm{inv}}(V)_{\mathbf{d},a} = p_{\mathrm{inv}}(V)_{\mathbf{d}}$.

Finally, we define $W^{(2)}$ to compute the sum in equation 57. In particular, for $X = [x_1,\dots,x_k] \in \mathbb{R}^{\tilde{D}n'\times k}$ we write $W^{(2)}X = [W_1^{(2)} x_1, \dots, W_k^{(2)} x_k]$, where $(W_b^{(2)} x_b)_a = \sum_{\mathbf{d}} x_{i_{\mathbf{d}},b}$. It can be seen that with these definitions of $W^{(2)}, W^{(1)}$, and q_{inv}, we have written p in the desired form.

D. Sign Equivariant Architecture Universality

In this section, we prove Proposition 3.2 on the universality of our proposed sign equivariant architectures, which we restate here:

Proposition 3.2. Functions of the form $v \mapsto v \odot \mathrm{MLP}(|v|)$ universally approximate continuous sign equivariant functions $f : \mathbb{R}^k \to \mathbb{R}^k$.

Compositions $f_2 \circ f_1$ of functions f_l as in equation 4 universally approximate continuous sign equivariant functions $f : \mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$.

We prove the two statements of the proposition in the next two subsections.

D.1. Universality for functions $\mathbb{R}^k \to \mathbb{R}^k$

Proof. Let $\mathcal{X} \subseteq \mathbb{R}^k$ be a compact set, let $\epsilon > 0$, and let $f_{\mathrm{target}} : \mathcal{X} \to \mathbb{R}^k$ be a continuous sign equivariant function that we wish to approximate within ϵ. Choose a sign equivariant polynomial p that approximates f_{target} to within $\epsilon/2$ on $\mathcal{X}$. By compactness, we can choose a finite bound $B > 0$ such that $|v_i| < B$ for all $v \in \mathcal{X}$.

By Proposition C.3, we can write $p(v)_l = v_l \cdot \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}$. By the universal approximation theorem for multilayer perceptrons, we can choose a MLP $: \mathcal{X} \to \mathbb{R}^k$ such that approximates $q(v) = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}$ up to $\epsilon/(2B)$. Note that $q(|v|) = q(v)$, so $v \mapsto \mathrm{MLP}(|v|)$ also approximates q within $\epsilon/(2B)$ accuracy.

Thus, for all $v \in \mathcal{X}$, we have that

$$|f(v)_i - p(v)_i| = |v_i \cdot \mathrm{MLP}(|v|)_i - v_i \cdot \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| \tag{59}$$

$$= |v_i||\mathrm{MLP}(|v|)_i - \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| \tag{60}$$

$$\leq B \cdot |\mathrm{MLP}(|v|)_i - \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| \tag{61}$$

$$< \epsilon/2, \tag{62}$$

so $\|f - p\|_\infty < \epsilon/2$ on $\mathcal{X}$ and we are done by the triangle inequality. □

D.2. Universality for functions $\mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$

Recall that each layer of our sign equivariant network from $\mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$ takes the form

$$f_l(V) = [W_1^{(l)} v_1, \ldots, W_k^{(l)} v_k] \odot \mathrm{SignNet}_l(V).$$

Proof. Let $\mathcal{X} \subseteq \mathbb{R}^{n \times k}$ be compact, and let $f_{\mathrm{target}} : \mathcal{X} \to \mathbb{R}^{n' \times k}$ be a continuous sign equivariant function that we wish to approximate. Since $\mathcal{X}$ is compact, we can choose a finite bound $B > 0$ such that $|V_{ij}| < B$ for all $V \in \mathcal{X}$. Let $p : \mathcal{X} \subseteq \mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$ be a sign equivariant polynomial that approximates f_{target} up to $\epsilon/2$ accuracy. Using Proposition C.4, we can write

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} V_{i_{\mathbf{d}},b} \cdot p_{\mathrm{inv}}(V)_{\mathbf{d}},$$

for some sign invariant polynomials $p_{\mathrm{inv}}(V)_{\mathbf{d}}$. We will have one network layer f_1 approximate the summands, and have the second network layer f_2 compute the sum.

First, we absorb the coefficients $\mathbf{W}_{\mathbf{d}}^{(a,b)}$ into the sign invariant part, by defining the sign invariant polynomial $q_{\mathrm{inv}}(V)_{\mathbf{d},a,b} =$

$\mathbf{W}_{\mathbf{d}}^{(a,b)} p_{\text{inv}}(V)_{\mathbf{d}}$, so we can write

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} V_{i_{\mathbf{d}},b} \cdot q_{\text{inv}}(V)_{\mathbf{d},a,b}.$$

Now, let $d_{\text{hidden}} \in \mathbb{N}$ denote the number of all possible $\mathbf{d}$ that appear in the sum, multiplied by n'. We define $f_1 : \mathcal{X} \to \mathbb{R}^{d_{\text{hidden}} \times k}$ as follows. As SignNet (Lim et al., 2023) universally approximates sign invariant functions on compact sets, we can let $\text{SignNet}_1 : \mathcal{X} \to \mathbb{R}^{d_{\text{hidden}} \times k}$ be a SignNet that approximates $q_{\text{inv}}(V)$ up to $\epsilon/(2B)$ accuracy, so

$$|\text{SignNet}_1(V)_{(\mathbf{d},a),b} - q_{\text{inv}}(V)_{\mathbf{d},a,b}| < \frac{\epsilon}{2B \cdot d_{\text{hidden}}}. \tag{63}$$

For $b \in [k]$, we also define the weight matrices $W_b^{(1)} \in \mathbb{R}^{d_{\text{hidden}} \times n}$ of the layer by letting the $(\mathbf{d}, a)$th row $(W_b^{(1)})_{(\mathbf{d},a),:}$ for any $a \in [n]$ only be nonzero in the $i_{\mathbf{d}}$th index, where it is equal to 1. Thus,

$$(W_b^{(1)} v_b)_{(\mathbf{d},a)} = V_{i_{\mathbf{d}},b}. \tag{64}$$

Hence, the first layer takes the form

$$f_1(V)_{(\mathbf{d},a),:} = \begin{bmatrix} V_{i_{\mathbf{d}},1} \cdot \text{SignNet}_1(V)_{(\mathbf{d},a),1} & \dots & V_{i_{\mathbf{d}},k} \cdot \text{SignNet}_1(V)_{(\mathbf{d},a),k} \end{bmatrix} \in \mathbb{R}^k. \tag{65}$$

Now, for the second layer, we let $\text{SignNet}(V)_{i,j} = 1$ for all $i \in [n], j \in [k]$, which can be represented exactly. Then for each column $b \in [k]$ we will define weight matrices $W_b^{(2)}$ such that $(W_b^{(2)})_{a,(\mathbf{d},i)} = 1$ if $a = i$ and is 0 otherwise. Then we can see that

$$f_2 \circ f_1(V)_{a,b} = \sum_{\mathbf{d}} V_{i_{\mathbf{d}},b} \cdot \text{SignNet}_1(V)_{(d,a),b}. \tag{66}$$

To see that this approximates the polynomial p, for any $V \in \mathcal{X}$ we can bound

$$\begin{aligned} |p(V)_{a,b} - f_2 \circ f_1(V)_{a,b}| &= \left| \sum_{\mathbf{d}} V_{i_{\mathbf{d}},b} \cdot \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| & (67) \\ &\leq \sum_{\mathbf{d}} |V_{i_{\mathbf{d}},b}| \left| \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| & (68) \\ &\leq B \sum_{\mathbf{d}} \left| \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| & (69) \\ &< B \sum_{\mathbf{d}} \frac{\epsilon}{2B d_{\text{hidden}}} & (70) \\ &\leq \frac{\epsilon}{2} & (71) \end{aligned}$$

By the triangle inequality, $f_2 \circ f_1$ is ϵ-close to f_{target}, so we are done.

□

E. Experimental Details

E.1. Miscellaneous Experimental Details

We ran the experiments on a HPC server with CPUs and GPUs. Each experiment was run on a single NVIDIA V100 GPU with 32GB memory. The runtimes for some of our experiments are included in the main paper. Our codes for our models and experiments will be open-sourced and permissively licensed.

E.2. Link Prediction in Nearly Synthetic Graphs

The base graphs H we generate are Erdös-Renyi (Erdos & Rényi, 1963) or Barabási-Albert (Barabási & Albert, 1999) graphs with 1000 nodes. We use NetworkX (Hagberg et al., 2008) to generate and process the graphs. The Erdös-Renyi graphs have edge probability $p = .05$ and the Barabási-Albert graphs have $m = 20$ new edges per new node. Let $V = [v_1, \dots, v_k]$

be Laplacian eigenvectors of the graph. We take $k = 16$ in these experiments. The unlearned decoder baseline simply takes the predicted probability of a link between i and j to be proportional to the dot product of the eigenvectors embeddings of node i and node j; this has no learnable parameters. In other words, the node embeddings z_i and z_j are taken to be $V_{i,:}$ and $V_{j,:}$ respectively, and the edge prediction is $z_i^\top z_j$. The learned decoder baseline takes the same z_i and z_j, but takes the edge prediction to be $\mathrm{MLP}(z_i \odot z_j)$. Every other method learns node embeddings z_i and z_j, and takes the edge prediction to be $z_i^\top z_j$.

Each model is restricted to around 25,000 learnable parameters (besides the Unlearned Decoder, which has no parameters). We train each method for 100 epochs with an Adam optimizer (Kingma & Ba, 2015) at a learning rate of .01. The train/validation/test split is 80%/10%/10%, and is chosen uniformly at random.

E.3. Details on n-body Simulations

We follow the experimental setting and build on the code of Puny et al. (2022) (no license as far as we can tell) for the n-body learning task. The code for generating the data stems from Kipf et al. (2018) (MIT License) and Fuchs et al. (2020) (MIT License). There are 3000 training trajectories, 2000 validation trajectories, and 2000 test trajectories. We modify the data generation code to apply to general dimensions $d > 3$. We do not change any of the scaling factors in doing so. For each dimension d, we use the same hyperparameters for both the frame averaging model and the sign equivariant model.

Appendix Citations

Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. In *Science*, pp. 509–512. American Association for the Advancement of Science, 1999.

Bronstein, M. M., Bruna, J., Cohen, T., and Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *preprint arXiv:2104.13478*, 2021.

Cohen, T. and Welling, M. Group equivariant convolutional networks. In *ICML*, pp. 2990–2999. PMLR, 2016.

Cohen, T. S. and Welling, M. Steerable CNNs. In *ICLR*, 2017.

Cotta, L., Bevilacqua, B., Ahmed, N., and Ribeiro, B. Causal lifting and link prediction. *preprint arXiv:2302.01198*, 2023.

Dym, N. and Gortler, S. J. Low dimensional invariant embeddings for universal geometric learning. *preprint arXiv:2205.02956*, 2022.

Dym, N. and Maron, H. On the universality of rotation equivariant point cloud networks. In *ICLR*, 2021.

Erdos, P. and Rényi, A. Asymmetric graphs. *Acta Math. Acad. Sci. Hungar*, 14(295-315):3, 1963.

Finzi, M., Welling, M., and Wilson, A. G. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *ICML*, pp. 3318–3328. PMLR, 2021.

Geisler, S., Li, Y., Mankowitz, D., Cemgil, A. T., Günnemann, S., and Paduraru, C. Transformers meet directed graphs. *preprint arXiv:2302.00049*, 2023.

Hagberg, A., Swart, P., and S Chult, D. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

He, X., Hooi, B., Laurent, T., Perold, A., LeCun, Y., and Bresson, X. A generalization of ViT/MLP-Mixer to graphs. *preprint arXiv:2212.13350*, 2022.

Kim, J., Nguyen, T. D., Min, S., Cho, S., Lee, M., Lee, H., and Hong, S. Pure transformers are powerful graph learners. In *NeurIPS*, 2022.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *ICLR*, 2015.

Kipf, T., Fetaya, E., Wang, K.-C., Welling, M., and Zemel, R. Neural relational inference for interacting systems. In *ICML*, pp. 2688–2697. PMLR, 2018.

Kondor, R. and Trivedi, S. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning*, pp. 2747–2755. PMLR, 2018.

Kreuzer, D., Beaini, D., Hamilton, W., Létourneau, V., and Tossou, P. Rethinking graph transformers with spectral attention. In *NeurIPS*, volume 34, 2021.

Lee, S.-L., Luo, Y.-L., Sagan, B. E., and Yeh, Y.-N. Eigenvector and eigenvalues of some special graphs. IV. Multilevel circulants. *International journal of quantum chemistry*, 41(1):105–116, 1992.

Li, P., Wang, Y., Wang, H., and Leskovec, J. Distance encoding: Design provably more powerful neural networks for graph representation learning. In *NeurIPS*, volume 33, pp. 4465–4478, 2020.

Maron, H., Ben-Hamu, H., Shamir, N., and Lipman, Y. Invariant and equivariant graph networks. In *ICLR*, volume 6, 2018.

Maron, H., Fetaya, E., Segol, N., and Lipman, Y. On the universality of invariant networks. In *ICML*, pp. 4363–4371. PMLR, 2019.

Mialon, G., Chen, D., Selosse, M., and Mairal, J. GraphiT: Encoding graph structure in transformers. In *preprint arXiv:2106.05667*, 2021.

Müller, L., Galkin, M., Morris, C., and Rampášek, L. Attending to graph transformers. *preprint arXiv:2302.04181*, 2023.

Murphy, M., Jegelka, S., Fraenkel, E., Kind, T., Healey, D., and Butler, T. Efficiently predicting high resolution mass spectra with graph neural networks. *preprint arXiv:2301.11419*, 2023.

Puny, O., Lim, D., Kiani, B. T., Maron, H., and Lipman, Y. Equivariant polynomials for graph neural networks. *preprint arXiv:2302.11556*, 2023.

Rampasek, L., Galkin, M., Dwivedi, V. P., Luu, A. T., Wolf, G., and Beaini, D. Recipe for a general, powerful, scalable graph transformer. In *NeurIPS*, 2022.

Ravanbakhsh, S., Schneider, J., and Poczos, B. Equivariance through parameter-sharing. In *ICML*, pp. 2892–2901. PMLR, 2017.

Segol, N. and Lipman, Y. On universal equivariant set networks. In *ICLR*, volume 7, 2019.

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., and Riley, P. Tensor field networks: Rotation-and translation-equivariant neural networks for 3D point clouds. In *NeurIPS Molecules and Materials Workshop*, 2018.

Villar, S., Hogg, D., Storey-Fisher, K., Yao, W., and Blum-Smith, B. Scalars are universal: Equivariant machine learning, structured like classical physics. In *NeurIPS*, volume 34, 2021.

Wang, H., Yin, H., Zhang, M., and Li, P. Equivariant and stable positional encoding for more powerful graph neural networks. In *ICLR*, volume 10, 2022.

Wood, J. and Shawe-Taylor, J. Representation theory and invariant neural networks. *Discrete applied mathematics*, 69(1-2): 33–60, 1996.

Yarotsky, D. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55(1):407–474, 2022.

You, J., Gomes-Selman, J. M., Ying, R., and Leskovec, J. Identity-aware graph neural networks. In *AAAI*, volume 35, pp. 10737–10745, 2021.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R., and Smola, A. J. Deep Sets. In *NeurIPS*, volume 30, pp. 3391–3401, 2017.

Zhang, M. and Chen, Y. Link prediction based on graph neural networks. In *NeurIPS*, volume 31, 2018.

Zhu, Z., Zhang, Z., Xhonneux, L.-P., and Tang, J. Neural bellman-ford networks: A general graph neural network framework for link prediction. In *NeurIPS*, volume 34, pp. 29476–29490, 2021.