Supplementary Information: Crystal Hypergraph Convolutional Networks

Alexander J. Heilman[1], Weiyi Gong[1], and Qimin Yan[1*]

[1]*Department of Physics, Northeastern University, Boston, MA 02115, USA*

S1 Motif Features: Structure Order Parameters & Continuous Symmetry Measures

The geometry of the motifs were incorporated as features composed of a concatenated list of structure order parameters and continuous symmetry measures (CSMs) for a set of common local environments.

Structure order parameters are coordinate system invariant measures of 3 dimensional structure that are designed to be close to one when a given structure is similar to some prototypical arrangement. Note that this isn't in general a true 'distance'-like measure to some shape as a CSM is, however. A CSM is essentially defined so that it may act as a 'distance' from some prototypical shape to some given structure. The list of order parameters included as motif features are those that implemented in existing pymatgen code and described in [1, 2].

S2 CHGConv

A specific implementation of a hypergraph convolutional operator in the hypergraph message passing framework is a generalization of CGConv implemented in pytorch geometric and based on CGCNN's convolutional operator defined in eq (5) of the original paper [11].

$$\begin{aligned} x_i^{t+1} &= \sum_{b_j} f(x_i^t, b_j, \mathrm{AGG}(\{x_j^t \in b_j\})) \\ &= \mathrm{BN}\Big[\sum_{b_j} \sigma\big(W_c \cdot [x_j \oplus b_j \oplus \mathrm{AGG}(\{x_j^t \in b_j\}]\big) \\ &\qquad \cdot S^+(W_f \cdot (x_j \oplus b_j \oplus \mathrm{AGG}(\{x_j^t \in b_j\})))\Big] \end{aligned}$$

For the model utilized in this work, the AGG function chosen was a combined set of component-wise maximum, minimum, average, and standard deviation, all with learnable attention weights. This combined use of multiple aggregation functions was inspired by a similar approach taken in the *ChemGNN* model [3].

S3 Tabulated Hyperparameters for Testing

For each convolutional structure, testing was done for a model with 3 convolutional layers. Each convolutional layer consists of back-to-back convolution from the smallest to the largest hyperedge type (for example two

*Corresponding Author: q.yan@northeastern.edu

bond & motif layers consist of a total sequence of bond, motif, bond and motif convolution). Stochastic gradient descent (SGD) was used as an optimizer through training with an initial learning rate of 0.01. A multi-step learning rate scheduler divided this learning rate by a factor of 10 at epoch 150, with training running for a total of 300 epochs. Hidden node features were of dimension 64 through all convolutional layers, and a hidden output layer of dimension 128 was used (similar to CGCNN's architecture). The loss functions utilized were Mean Squared Error (MSE, for regression tasks) and cross entropy (for classification tasks). Accuracy is then reported in Mean Absolute Error (MAE) for regression tasks and area under curve (AUC) for classification tasks. For convenience, these parameters are summarized in table S1.

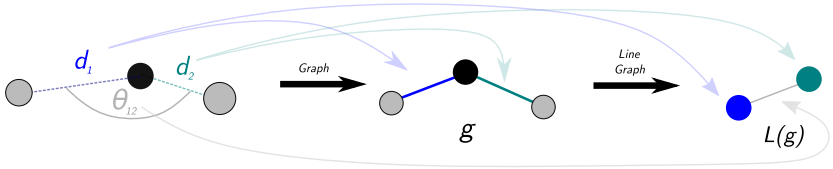
Hyperparameter	**Value**
Node Hidden Feature Dimension	64
Post-Convolution Linear Width	128
Number of Convolutional Layers	3
Number of Epochs	300
Batch-size	16 ($< 10,000$ Samples) *or* 32 ($< 20,000$) *or* 64
Optimizer	SGD
Learning Rate (Epoch <150)	0.01
Learning Rate (Epoch >150)	0.001
Loss Function	MSE (Regression) *or* Cross Entropy (Class.)

Table S1: Tabulated hyperparameters used for model with reported results.

Results reported were averaged over the 5 pre-fixed MatBench [4] folds for nested cross-validation. These datasets were divided into 80% for training and 20% for test for each fold, with a further 10% of the training subset being withheld from training and used as an indicative validation set, where the best performance on this dataset was used to select the model applied to the test set.

S4 Comparison to Line Graph

A more usual approach for the incorporation of bond angle information is via the construction of a line graph, as in [5, 6].



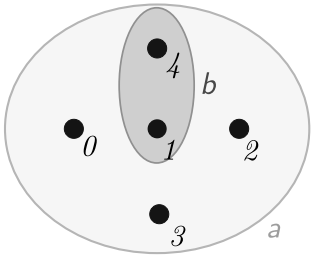
These models generally first update the edge features of the crystal graph $\mathcal{G}$ by first applying some graph convolutional operator to the line graph $L(\mathcal{G})$ with angles encoded in $L(\mathcal{G})$'s initial edge features.

Our argument against such representation schemes here is that the order of messages grows combinatorically for derived line graphs as $\mathcal{O}(nm^2)$, where n is the number of nodes and m is the average number of edges per node in $\mathcal{G}$.

Here, we incorporate a similar level of higher-order geometrical structure instead in a local environment, or 'motif', hyperedge (defined below). Note that these include only an extra number of messages on the order $\mathcal{O}(mn)$ if each node in a motif gets a message, or on the order $\mathcal{O}(n)$ if only center nodes are updated by their own motif hyperedges.

S5 Hyperedge Index

Hypergraphs are treated as a set of node feactures x, hyperedge features h, and hyperedge indices I. The hyperedge index is, computationally, treated as a $[2, nm]$ dimensional vector (where m is the number of hyperedges and n is the avereage number of nodes contained in any hyperedge). The first index is the node contained and the second index is the containing hyperedge (as in [7]). This scheme is depicted in an example in Figure S1



Hyperedge Index: [[0,1,2,3,4,1,4],
[a,a,a,a,a,b,b]]

Figure S1: Hyperedge index example for a specific hypergraph.

S6 Message Number Scaling for Different Hyperedge Types

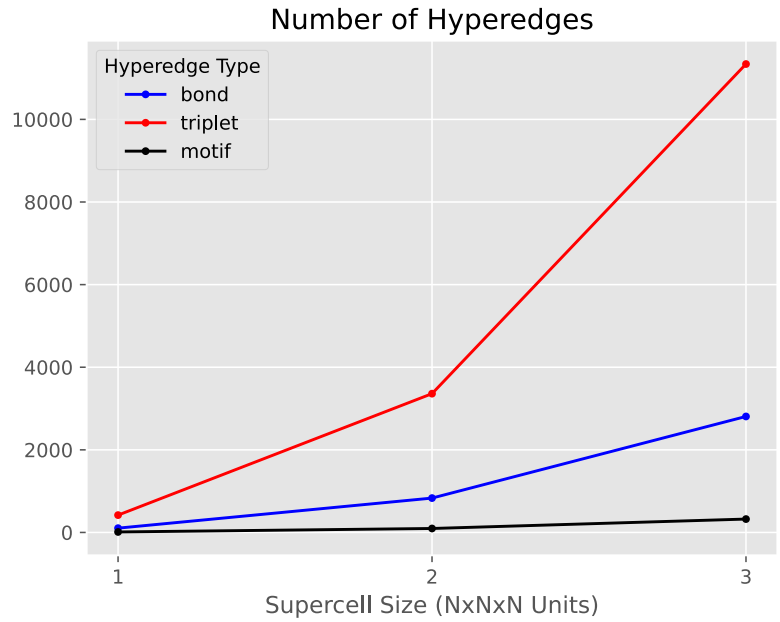
The rationale for including motifs in lieu of triplets (as in line graphs) is most poignant when considering material systems of increasing unit-cell size. To make this point most clear, here we consider the number of hyperedges required for the three different types of hyperedges considered in this work for different sized supercells of Manganese Oxide (MnO_2) in Figure S2.

Here, the number of triplets is shown to grow quadratically with respect to the number of edges per node, whereas the number of bonds grows quadratically with respect to the number of atoms, and the number of motifs grows linearly. As such, additional geometric resolution may be afforded by relatively few motif hyperedges, as compared to triplet-based constructions.

S7 Case Study on Discrimination of Similar Environments

To demonstrate the importance of motif-level hyperedges in applications to material systems, the node embeddings of two compositionally-similar, but structurally distinct materials are considered here: Trigonal (T-phase) and Hexagonal (H-phase) MoS_2. To demonstrate the effectiveness of the motif models in sooner recognizing structural differences, the difference in the node embeddings of these two models are compared through 300 epochs of training for models with 1 convolutional layer of either bond-, motif-, or triplet-level hyperedges. Node representation difference was calculated by subtracting the normalized dot-product of the central atom's node embeddings from one. This node representation difference through training is shown in Figure S3.

While bond-only models could slightly distinguish the node embeddings of central atoms between these different, but similar, structures, the motif-based model captured the structural differences earliest (that is,



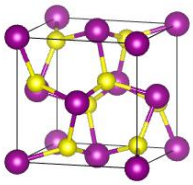
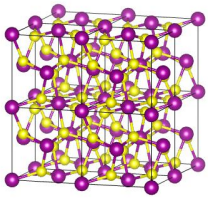
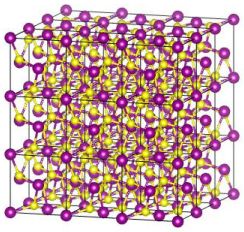
	Conventional Cell	$2 \times 2 \times 2$ Super Cell	$3 \times 3 \times 3$ Super Cell
			
Bonds:	104	832	2808
Triplets:	420	3360	11340
Motifs:	12	96	324

Figure S2: Figure tabulating the number of hyperedges of each type for increasing supercell size. The example material here is Manganese Oxide (MnO_2).

the one-layer models of motifs always show the greatest difference). Specifically, the motif-only model had a representation difference of 0.004 after 1 epoch of training, whereas the triplet and bond-only models had an initial difference of 0.0005 and 0.00025, respectively. After 300 epochs, the motif-only model still had the largest difference of 0.075, while the triplet model's difference had increased to 0.008 and the bond-only model had remained similar with a difference of 0.00029. This demonstrates that the motif-only model had the greatest discrimination between the two similar environments, a difference relevant to the band gap target since T-phase MoS_2 is metallic (with a band gap of zero), and H-phase MoS_2 is not (with a band gap of 1.2 eV). However, it should be noted this metric is unreliable, insofar that neural networks are essentially black boxes and hence sheer differences in node embeddings are hard to translate directly to model performance.

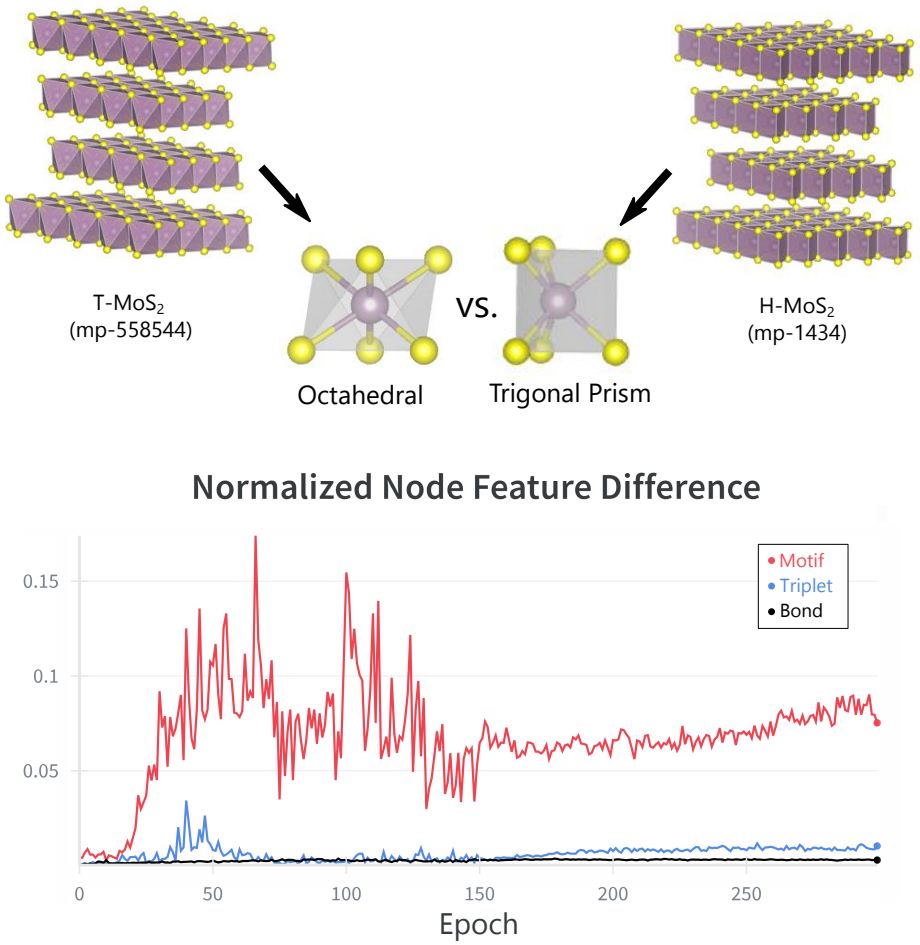


Figure S3: Normalized dot product difference between analogous H-phase and T-phase MoS_2 Molybdenum node representations across 300 epochs of 1 layer of either bond-, motif-, or triplet-only convolution. Here, red is the difference across epochs for the motif-only model; blue is for triplet-only model; and black represents the difference for the bond-only model. Note that the identification numbers below the material labels are the Materials Project [8] identification numbers.

S8 Comparison of Order Parameters and Continuous Symmetry Measures as Motif Features

To compare the importance of motif features, the data set that most benefited from motif features was tested both with and without the two primary types of motif features considered in this work, namely: 35 Local Structure Order Parameters (LSOPs) as defined in [1, 2]; and Continuous Symmetry Measures [9, 10] for 59 common coordination polyhedra. Motif convolutional models consisting of three motif hypergraph convolutional layers were tested on this bulk moduli dataset for 300 epochs with the following feature sets: no motif features, only CSMs, only LSOPs, or both feature sets (concatenated). For all tests, the batch size was 64. Stochastic Gradient Descent (SGD) was used as the optimizer with an initial learning rate of 0.1, and a reduced learning rate of 0.01 being applied at epoch 150. These results are tabulated in Table S2.

CSM-only features performed best on the validation set, suggesting they offer better generalizatbility

Motif Features	**Test MAE** $\mathrm{Log}_{10}(G_{vrh})$
None	0.1038
CSMs	0.0905
LSOPs	0.0939
Both	0.0917

Table S2: Table showing validation results for a model with 1 motif hypergraph convolutional layer on the bulk moduli target set after 300 epochs of training. The shear moduli MatBench [4] dataset is used with a 80% training / 20% test split of the total data set. The training subset was further split into a 90% training / 10% validation set, with the best performing model on the withheld validation set then used on the test set for reported results.

as a feature. This may be due to the fact that CSM-based features result in sparse (similar to one-hot) encodings, whereas LSOP features tend more towards scalar-valued features. That is, while both CSMs and LSOPs are scalar-valued, a large subset of incommensurate ideal shapes used in CSM calculations results in many zero entries, whereas LSOPs may generally be computed regardless and just return smaller scalar values, so that the sparse CSM features may more adequately distinguish environments through the model. Consequently, only CSM features are used in the results reported in the main text.

S9 Performance on Matbench Folds

Results tabulated in the main text are averaged over 5 folds of nested cross-validation, with test indexes supplied by MatBench [4]. Accordingly, the uncertainty in test performance Δx is calculated as:

$$\Delta x = \frac{\sigma}{\sqrt{N}} \tag{1}$$

with σ the standard deviation of the values and N the number of runs (here, 5 for the 5 folds). Furthermore, the test set performance for each fold is tabulated below. Furthermore, in the following, $\bar{x}$ denotes the average of the values in the column above. These results are shown in Tables S3-S11 below. Note that in the following the bond-only results reported are CGCNN's [11] reported MatBench results since our code is implemented directly into CGCNN's publicly available source code (https://github.com/txie-93/cgcnn).

Perovskite - Formation Energy Test MAE (meV/atom)			
Fold	Bond-only	Bond & Motif	Bond & Triplet
1	45.6	42.7	42.2
2	46.2	43.3	42.7
3	44.8	42.0	43.1
4	45.4	40.4	42.3
5	44.2	43.0	42.9
$\bar{x}$	45.2	42.3	42.6
Δx	0.3	0.5	0.2

Table S3: Performance across the 5 MatBench test folds on 18,928 materials of the perovskite formation energy target set.

$\text{Log}_{10}(\text{K}_{vrh})$ Test MAE (Log_{10}GPa)			
Fold	Bond-only	Bond & Motif	Bond & Triplet
1	0.0702	0.0627	0.0649
2	0.0722	0.0649	0.0669
3	0.0665	0.0613	0.0672
4	0.0748	0.0667	0.0680
5	0.0724	0.0623	0.0661
$\bar{x}$	0.0712	0.0636	0.0666
Δx	1.2×10^{-3}	0.9×10^{-3}	0.5×10^{-3}

Table S4: Performance across the 5 MatBench test folds on 10,987 materials of the bulk moduli target set.

$\text{Log}_{10}(\text{G}_{vrh})$ Test MAE (Log_{10}GPa)			
Fold	Bond-only	Bond & Motif	Bond & Triplet
1	0.0870	0.0784	0.0840
2	0.0899	0.0815	0.0866
3	0.0887	0.0817	0.0853
4	0.0902	0.0797	0.0825
5	0.0918	0.0813	0.0844
$\bar{x}$	0.0895	0.0805	0.0846
Δx	7×10^{-4}	6×10^{-4}	6×10^{-4}

Table S5: Performance across the 5 MatBench test folds on 10,987 materials of the shear moduli target set.

Highest Phonon Peak ω Test MAE (cm^{-1})			
Fold	Bond-only	Bond & Motif	Bond & Triplet
1	81.2	59.7	79.8
2	45.1	43.8	40.8
3	54.3	55.3	63.7
4	56.6	55.9	56.8
5	51.7	50.0	62.4
$\bar{x}$	57.8	52.9	60.7
Δx	5.5	2.5	5.6

Table S6: Performance across the 5 MatBench test folds on 1,265 materials of the phonon peak target set.

Dielectric Constant Test MAE			
Fold	Bond-only	Bond & Motif	Bond & Triplet
1	0.470	0.343	0.353
2	0.572	0.419	0.380
3	0.730	0.533	0.582
4	0.611	0.447	0.476
5	0.610	0.415	0.407
$\bar{x}$	0.599	0.431	0.440
Δx	0.037	0.027	0.037

Table S7: Performance across the 5 MatBench test folds on 4,764 materials of the dielectric constant target set.

Exfoliation Energy (Jdft2d) Test MAE (meV/Atom)			
Fold	Bond-only	Bond & Motif	Bond & Triplet
1	32.5	48.4	38.5
2	51.1	44.8	57.4
3	69.4	63.1	77.2
4	42.7	30.4	64.6
5	48.4	47.9	56.3
$\bar{x}$	48.8	46.9	58.8
Δx	5.4	4.7	5.6

Table S8: Performance across the 5 MatBench test folds on the exfoliation energy target set for a set of 626 2D materials (labelled Jdft2d by MatBench).

MP - Band Gap Test MAE (meV)		
Fold	Bond-only	Bond & Motif
1	0.298	0.239
2	0.294	0.229
3	0.296	0.226
4	0.295	0.222
5	0.304	0.239
$\bar{x}$	0.297	0.231
Δx	0.002	0.003

Table S9: Performance across the 5 MatBench test folds on 106,113 materials of the Materials Project band gap target set.

MP - Formation Energy Test MAE (meV/atom)		
Fold	Bond-only	Bond & Motif
1	34.0	34.7
2	34.0	40.2
3	32.8	36.9
4	33.2	49.8
5	34.6	37.0
$\bar{x}$	33.7	39.7
Δx	0.3	2.4

Table S10: Performance across the 5 MatBench test folds on 132,752 materials of the Materials Project formation energy target set.

MP - Metalicity Test AUC		
Fold	Bond-only	Bond & Motif
1	.959	.960
2	.945	.957
3	.964	.958
4	.948	.960
5	.951	.955
$\bar{x}$	0.953	0.958
Δx	3×10^{-3}	1×10^{-3}

Table S11: Performance across the 5 MatBench test folds on 106,113 materials of the Materials Project metallicity target set.

References

[1] N. E. R. Zimmermann, M. K. Horton, A. Jain, and M. Haranczyk, *Front. Mater.* **4**, 34 (2017).

[2] N. E. Zimmermann and A. Jain, *RSC Adv.* **10**, 6063–6081 (2020).

[3] C. Chen, E. Xu, D. Yang, C. Yan, T. Wei, H. Chen, Y. Wei, and M. Chen, *Neural Comput. Apl.* **37**, 3287–3301 (2024).

[4] A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, *npj Comput. Mater.* **6**, 138 (2020).

[5] K. Choudhary and B. DeCost, *npj Comput. Mater.* **7**, 1–8 (2021).

[6] C. Chen and S. P. Ong, *Nat. Comput. Sci.* **2**, 718–728 (2022).

[7] S. Bai, F. Zhang, and P. H. Torr, *Pattern Recogn.* **110**, 107637 (2021).

[8] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, *APL Mater.* **1**, 011002 (2013).

[9] M. Pinsky and D. Avnir, *Inorg. Chem.* **37**, 5575–5582 (1998).

[10] J Lima-de Faria, E Hellner, F Liebau, E Makovicky, and E Parthé, *Acta Crystallogr. A* **46**, 1–11 (1990).

[11] T. Xie and J. C. Grossman, *PRL* **120**, 145301 (2018).