Multiscale Graph Attention Neural Networks for Mapping Materials and Molecules beyond Short-Range Interatomic Correlations

Yuanbin Liu

Tsinghua University

Xin Liu

Harbin Institute of Technology

Bingyang Cao (✉ caoby@tsinghua.edu.cn)

Tsinghua University <https://orcid.org/0000-0003-3588-972X>

Article

Keywords: neural networks, machine learning, unified framework, metal-organic frameworks

Posted Date: November 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1064739/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Multiscale Graph Attention Neural Networks for Mapping Materials and Molecules beyond Short-Range Interatomic Correlations

Yuanbin Liu¹, Xin Liu², Bingyang Cao^{1,*}

¹*Key Laboratory for Thermal Science and Power Engineering of Ministry of Education, Department of Engineering Mechanics, Tsinghua University, Beijing 100084, China*

²*School of Chemical Engineering and Advanced Materials, The University of Adelaide, Adelaide, SA 5005, Australia*

[*caoby@tsinghua.edu.cn](mailto:caoby@tsinghua.edu.cn)

Abstract: Bringing advances of machine learning to chemical science is leading to a revolutionary change in the way of accelerating materials discovery and atomic-scale simulations. Currently, most successful machine learning schemes can be largely traced to the use of localized atomic environments in the structural representation of materials and molecules. However, this may undermine the reliability of machine learning models for mapping complex systems and describing long-range physical effects because of the lack of non-local correlations between atoms. To overcome such limitations, here we report a unified framework named the multiscale graph attention neural network to map materials and molecules into a generalizable and interpretable representation which combines local and non-local information of atomic environments from multiple scales. As an exemplary study, our model is applied to predict electronic structure properties of one class of technologically important reticular materials, i.e., metal-organic frameworks (MOFs) which have notable diversity in compositions and structures. Our model is trained on density functional theory calculated datasets and the results show that it achieves the state-of-the-art performance for such a challenging task. The clustering analysis further demonstrates that our model leads to a high-level identification of MOFs with spatial and chemical resolution, which would be capable of revealing new insights into complex systems and efficiently guiding the search for reticular materials with desired properties.

INTRODUCTION

The past few years have witnessed a surge of interests in applying machine learning (ML) technologies to power many aspects of both computational chemistry and materials science¹⁻³. For example, ML techniques open new avenues in the construction of complicated potential energy surface from quantum-mechanical data in an automated fashion to tackle the long-standing computational challenges (e.g., the realistic modeling of chemical reactions or complex materials and interfaces) that are inaccessible to either poorly transferrable empirical force fields or computationally demanding *ab initio* methods⁴⁻¹⁰. What is more, ML approaches are revolutionizing the discovery and design of materials and molecules at an astounding rate through direct *in silico* screening and statistical analysis of massive chemical data sets¹¹⁻²¹.

The representation of materials and molecules is a crucial ingredient to construct effective ML models dedicated to statistical property regression, to clustering of chemical structures, or to visualization of material phase space in a low-dimensional manifold. Conventional ML models adopt hand-crafted descriptors that encode the raw information about atomic systems (such as the chemical nature and coordinates of each atom) into a suitable representation with physical symmetry, such as widely used smooth overlap of atomic positions (SOAP)^{5, 22}, Coulomb matrix²³, atom-centered symmetry function^{4, 24}, and composition-based features²⁵. Arguably, the design of effective descriptors usually requires both considerable domain expertise and human efforts, which proves challenging for an immense number of materials with complex structures and many compositions.

Recently, tremendous attention has been received in the development of deep learning models to automatically discover the flexible representations of materials and molecules with minimal human intervention. Notably, deep learning allows raw data directly as inputs from which complex and abstract material representations can be learned by a series of hierarchically nested neural networks²⁶. To date, a number of deep learning models have been proposed for concerning problems especially in both material and chemical science, including deep tensor neural network (DTNN)²⁷, ANI-1²⁸, crystal graph convolutional neural network (CGCNN)²⁹, SchNet³⁰, PhysNet³¹, and AIMNet^{32, 33}. Those deep models have shown strong and flexible capability to represent complex systems (such as protein-like compounds³¹ and drug-like molecules¹¹) and generally outperform conventional ML methods for predicting the various

quantum-chemical properties of organic small molecules, crystals, disordered materials, and surfaces³⁴⁻³⁸.

Aside from powerful representation learning, the idea of locality undoubtedly lays a solid basis for current state-of-the-art (SOTA) ML-based potentials or property regression schemes. Locality, supported by the principle of electronic nearsightedness^{39, 40}, is associated with the description for atom-centered short-range chemical environments to infer complex many-body interaction and renders ML models interpretable, scalable, and robust for extensive properties^{41, 42}. In the context of locality, most of current SOTA models only build global representations as a collection of atom-centered local environments and neglect long-range interatomic correlations. However, this could significantly undermine the reliability of ML schemes when long-range interactions beyond the cutoff radius (such as electrostatics and van der Waals dispersion) dominate the properties of systems like ionic solids and electrolyte solutions⁴³⁻⁴⁵. Moreover, only encoding local information may lose global shape characteristics and prominently weaken pattern recognition for complex systems with widely structural and configurational diversity^{35, 46, 47}. Indeed, there is a great and increasing demand for ML to accelerate the discovery of complex materials. One class of representative materials is metal-organic frameworks (MOFs) that have great potential in many applications, such as gas storage and separation, sensors, thermoelectric, catalysis, and photovoltaics⁴⁸⁻⁵¹. Notably, over 80,000 nonporous MOFs have been synthesized over the past decade by assembling the organic linkers and metal clusters^{50, 52}, but they are just a small part of the myriads of possible structural motifs of MOFs that can be yielded⁵³. This results in an urgent need for ML in this area. However, the overwhelming chemical space plus the large number of atoms in MOF structures makes it difficult to obtain a high-level global representation only incorporating short-range interatomic correlations.

In this work, we introduce a unified multiscale graph attention neural network (MGANN) architecture that allows to capture local and non-local features of atomic environments, aiming to provide a deep and high-level representation for complex materials and molecules. We evaluate our model on a recent quantum-chemical database⁴⁷ of MOFs for predicting their electronic bandgaps. This exemplary study demonstrates that our model achieves the similar accuracy with respect to density-functional theory (DFT) calculations and outperforms the prior SOTA models, manifesting its general applicability. Finally, we illustrate how the learned information-rich

representations can be used in the high-fidelity chemical clustering and sharply narrow down the candidate space for fast searching of desired materials.

RESULTS

Multiscale graph attention neural networks. Figure 1 depicts a unified MGANN architecture proposed in this work. Inspired by the interpretability, generalizability, and remarkable performance of deep graph networks in predicting material properties^{34, 36, 38, 54, 55}, MGANN receives atomistic structures of materials through a graph-based descriptor where the atoms and the bonds that connect the atoms are regarded as the nodes and the edges in the graph, respectively. Following a routine protocol in graph-based models^{29, 38}, the initial node attributes are F -dimensional one-hot encodings of the chemical properties of elements, independent of atomic environments for now. The edge attributes between nodes i and j are encoded as a set of Gaussian-expanded distances $\mathbf{g}_{ij} = [g_{ij}^1, g_{ij}^2, \dots, g_{ij}^K]$ whose component value takes the form

$$g_{ij}^k = \exp(-\gamma(r_{ij} - \mu_k)^2) \quad (1)$$

where r_{ij} is the interaction distances between atoms i and j ; γ is an adjustable parameter specifying the width of Gaussian basis; $\mu_k = R_c k / (K - 1)$ for $k = \{0, 1, 2, \dots, n_{bf} - 1\}$ and R_c is the cutoff radius and K is the number of bond features. Herein, an undirected graph $\mathcal{G}_i = (\Gamma_i, \mathbf{g}_i)$ is ready for describing the local configuration of atom i , where $\Gamma_i = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{N_{li}}\} \subseteq \mathbb{R}^{F \times N_{li}}$ and $\mathbf{g}_i = \{\mathbf{g}_{i1}, \mathbf{g}_{i2}, \dots, \mathbf{g}_{iN_{li}}\} \subseteq \mathbb{R}^{N_{li} \times K}$ are a collection of the node and edge attributes in the local graph; N_{li} denotes the number of all neighbors of the i -th atom within a certain cutoff radius R_c or usually a fixed number of the nearest neighbors to save computer memory especially for those unit cells with large size. The whole system is naturally described by an undirect multigraph consisting of all local graphs.

Fig. 1. Illustration of the multiscale graph attention neural networks. (a) Architecture of overall MGANN model. MGANN receives atomistic structures of materials through graph-based descriptors where the atoms and the bonds that connect the atoms are regarded as the nodes and the edges in the graph, respectively. After the raw descriptors flow through a stack of the graph blocks and self-attention blocks, the local and nonlocal information will be extracted hierarchically on a large scale, and this eventually arrives at a high-level and global embedding of atomic structures. (b) Architecture of the GCNN blocks. (c) Architecture of the self-attention blocks. (d) Schematic of mapping molecules by the MGANN model. (e) Evaluation of the attention weights of any local atomic environmental (LAE) against itself and each of all others in the whole systems and over a long distance by the scaled dot-product attention operators.

A graph convolutional neural network module is built to upgrade the atomic embeddings by passing messages from neighbors and bonds. Distinct from previous SOTA deep graph networks like CGCNN²⁹ and SchNet³⁰, we exploit new bond convolution (BondConv) operations to directly extract the interaction features from all the bonds emanating from each atom. The updated atomic embeddings are outputted by a channel-wise symmetric aggregation operation, i.e., max pooling to

remain permutational invariance, and takes the form in the first convolutional layer

$$\mathbf{v}_i^{(l+1)} = \max_{j:(i,j) \in \mathcal{G}_i} \text{ReLU}(\boldsymbol{\varphi}_m \cdot \mathbf{v}_i^{(l)} + \boldsymbol{\theta}_m \cdot \mathbf{v}_j^{(l)} + \boldsymbol{\phi}_m \cdot \mathbf{g}_i^{(l)}) \quad (2)$$

where $(\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_M, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M, \boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_M)$ are the learnable weight parameters of atomic features and interactions; l denotes the current layer, here $l = 0$. In the sequent several convolutional layers, we change the BondConv to a simpler but effective form

$$\mathbf{v}_i^{(l+1)} = \max_{j:(i,j) \in \mathcal{G}_i} \text{ReLU}(\boldsymbol{\varphi}'_m \cdot \mathbf{v}_i^{(l)} + \boldsymbol{\theta}'_m \cdot (\mathbf{v}_j^{(l)} - \mathbf{v}_i^{(l)})). \quad (3)$$

Both Eqs.2 and 3 can be implemented by a shared multilayer perceptron, guaranteeing permutational invariance to the ordering of neighbors. Similar convolutional operations have been successfully used in the visual tasks on 3D point clouds⁵⁶, but not yet to atomic structures. Finally, the graphs are updated through each convolutional layer of the networks by increasingly embedding more information of local environments into atomic features. Meanwhile, besides aforementioned permutational invariance, the outputs are also strictly invariant to translation and rotation because atom-centered descriptors and only pairwise distances are used in the networks.

To address the losing the long-range information correlations between atoms in previous local graph representations, for the first time, we introduced the self-attention mechanism to our architecture. The self-attention mechanism was originally proposed and applied in the emerging Transformer architectures for boosting the performance of neural machine translation and the speed of the model training⁵⁷. The biggest benefit of self-attention comes from the fact that it allows processing the word sequences in parallel and make the modeling of the dependencies between words without regard to their distance in the input sequences. We notice that a few studies⁵⁸⁻⁶¹ have just recently introduced self-attention to mapping the space of chemical reactions from text-based representations, namely SMILES^{62, 63}, but none involves mapping atomic configurations to the best of our knowledge.

We now illustrate how self-attention is implemented in our model. Let the node-level embedding sets outputted from the graph networks be $\mathbf{G}_{em} \in \mathbb{R}^{N \times d_g}$ with the number of atoms N in a system and the feature dimensionality d_g . Regardless of the graph structure, the node set in the graph can be served as a N -component and out-of-order sequence. To compute self-attention, three matrices — *query* $\mathbf{Q}$, *key* $\mathbf{K}$, and *value* $\mathbf{V}$ as defined in the original literature⁵⁷ — need to be

created by linear transformations of the input features $\mathbf{G}_{em}$ as follows:

$$(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{G}_{em} \cdot (\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V) \quad (4)$$

where $\mathbf{W}_Q \in \mathbb{R}^{d_g \times d_a}$, $\mathbf{W}_K \in \mathbb{R}^{d_g \times d_a}$, and $\mathbf{W}_V \in \mathbb{R}^{d_g \times d_g}$ indicates the shared learnable linear transformation matrices, and d_a is the dimension of the query or key vector. With the dot product between the query and key matrices, we can evaluate the attention weights of any local atomic environment against itself and each of all others in the whole systems and over a long distance. The attention weight matrix $\mathbf{A}$ takes the form:

$$\mathbf{A} = \mathbf{Q} \cdot \mathbf{K}^T. \quad (5)$$

The attention weights determine how relevant the information of a certain local atomic environment is to that of other local and nonlocal ones, from which the long correlation of information can be built. To make gradients more stable and all attention weights positive, the attention weights are further scaled by a factor $1/\sqrt{d_a}$ and normalized by a softmax operation:

$$\tilde{\mathbf{A}} = \text{softmax}\left(\frac{\mathbf{A}}{\sqrt{d_a}}\right). \quad (6)$$

The outputs $\mathbf{F}_{sa}$ of the self-attention layer are expressed by summing up the weighted value vectors:

$$\mathbf{F}_{sa} = \tilde{\mathbf{A}} \cdot \mathbf{V}. \quad (7)$$

Here, multiplying each value vector by the softmax weights is to automatically drown out irrelevant information between local atomic environments and keep intact those that are worth attending. As all operators in self-attention are independent to the order and size of inputs, this endows our model with strictly permutational invariance and scalability. The self-attention module is shown in Fig. 1c. In fact, there are many available variants of self-attention that can be used to enhance the model. An alternative is to employ an evolved self-attention module — offset-attention — to replace the original one, which is inspired by the benefits of the Laplacian operator used in graph convolution networks⁶⁴. One can refer to our recent work for more details on offset-attention⁶⁵.

After the raw descriptors flow through a stack of the graph and self-attention layers, the local

and nonlocal information will be extracted hierarchically on a large scale, and this eventually arrives at a high-level and global embedding of atomic structures. Finally, the complex mappings from atomic structures to material properties are established by two fully connected hidden layers. All weights and other learnable parameters in the networks are iteratively updated by using mini-batch stochastic gradient descent and minimizing the difference between the predicted properties and the reference data computed by DFT.

Quantum MOF (QMOF) database. MOFs are a class of promising porous materials, and the fascinating aspects of MOFs are their synthetic versatility, chemical tunability, and stability. The isoreticular principle enables the size of MOFs to vary in a wide range without changing their underlying topology⁶⁶. For instance, the tunable pore aperture and surface area of MOFs could range from $< 10 \text{ \AA}$ to $\sim 100 \text{ \AA}$ and 1000 to 10,000 m^2/g ^{49, 67}, respectively. It allows one to fine-tune the structures of these materials with respect to selectivity and activity⁶⁸. To date, more than ten thousand kinds of MOFs have been synthesized by assembling the organic linkers (benzene-1,4-dicarboxylate, 2,5-dihydroxybenzene-1,4-dicarboxylate, biphenyl-4,4'-dicarboxylate, *etc.*) and metal clusters (Mn, Fe, Co, Cu, Zn, Ni *etc.*)^{52, 53}.

The QMOF database is a collective MOF subset of the Cambridge Structural Database⁵² and the 2019 CoRE MOF database⁶⁹. All crystal structures in the QMOF database are experimentally synthesized and fully relaxed at the DFT+PBE-D3(BJ) level⁷⁰⁻⁷². Compared with the previous MOF databases, such as the OQMD⁷³ and the CoRE MOF database⁷⁴, the QMOF database particularly calculates an important electrical structure property of MOFs — bandgap, E_g in eV. The bandgap is an excellent indicator for classifying MOFs into metal or semiconductor. Given that the majority of MOFs are electrical insulators, it is essential to identify the metallic MOFs or those with low bandgaps for expanding applications of MOFs into (opt-)electronic devices and revealing novel quantum-chemical insight into MOFs^{49, 75, 76}. The periodic table with color (Fig. 2a) shows 78 chemical elements covered in the QMOF database. The violin plots (Fig. 2b) further illustrate the statistical distributions of sizes per primitive unit cell in the QMOF databases. The multiple chemical elements (Fig. 2a) associated with the diverse structures (Fig. 2b) show the complexity of the MOF chemistry and make the QMOF database an excellent modeling target to assess the generality of our MGANN model.

Fig. 2. Structural properties in the QMOF database. (a) Periodic table with color showing 78 chemical elements covered in the QMOF database. The yellow, dark gray, and blue ones indicate the nonmetallic, metal, and lanthanide/actinide elements, respectively. (b) Violin plots of the statistical distributions of sizes per primitive unit cell in the QMOF databases. Separate distributions are shown for the three evolving versions of the QMOF databases: QMOF-1 (13,058 structures), QMOF-2 (14,482 structures), and QMOF-3 (18,333 structures), respectively. The (minimum, maximum, average) values of unit cell sizes in the three databases are (17, 150, 85), (17, 300, 95), and (17, 500, 103), respectively.

Learning bandgaps of MOFs. We now move to evaluate the performance of the MGANN model on the QMOF benchmark set and make a comprehensive comparison with other common ML models. Here, we simply divided the benchmark models into two categories: classical machine learning (CML) models with hand-crafted descriptors and deep learning (DL) models with learnable representations. The original work⁴⁷ of the QMOF database provides the benchmarks for

a DL model (i.e., CGCNN) and five CML models in terms of the QMOF-2 dataset. The five CML models are constructed with the same kernel ridge regression method but different descriptors, i.e., Sine Coulomb matrix (SineCM)⁷⁷, “Stoichiometric-45” (SM-45)⁷⁸, “Stoichiometric-120” (SM-120)²⁵, Orbital field matrix (OFM)⁷⁹, and SOAP. Additionally, SchNet is a representative graph-based DL model and is also involved as a comparison benchmark. For direct comparison with the benchmarks from the QMOF database, we also train MGANN and SchNet on the same QMOF-2 dataset, with the mean absolute error (MAE) and Spearman rank-order correlation coefficient (ρ) as the joint metrics to quantitatively gauge the performance of different models. The dataset is randomly split into 80% for training, 10% for validation, and 10% for testing. The random splitting is repeated five times for five parallel runs over which the statistics of MAEs and ρ on the testing sets are obtained. Notably, SchNet is trained by using the previous hyperparameters particularly optimized for MOFs³⁵, while such a fine exploration in the hyperparameter space is not carried out for MGANN.

As shown in Fig. 3a, on the one hand, SOAP achieves the best performance among the CML models. It indicates that the descriptors sensitive to atomic structures and chemical elements are essential to improve the model performance. In fact, SOAP has been proved to perform equally well with other SOTA models in building machine-learning potentials for systems containing a few elements⁸⁰. On the other hand, the DL models (CGCNN, SchNet, MGANN) substantially outperform all CML models in predicting the properties of structure-complex and element-diverse MOFs. The three DL models simply take raw molecular graphs consisting of atomic attributes and Gaussian basis as inputs, while they achieve the SOTA performance. It highlights the strength of the DL models to learn complex and abstract molecular representations out of simple ones, which can significantly reduce the human effort in designing suitable descriptors. Remarkably, our MGANN model outperforms the prior SOTA DL models, i.e., CGCNN and SchNet. It shows that our proposed graph convolutional operators are effective to capture the configurational information of local atomic environments and embedding atoms in a suitable way. In addition, the introduction of self-attention mechanism leads to a strong correlation between atoms over a long distance and incorporates more non-local information into global representations, thus boosting the accuracy of the MGANN model. Fig. 3b provides a correlation plot for predicted quantities of the MGANN model. We further train and evaluate the MGANN on the QMOF-3 dataset which

contains larger unit cells with up to 500 atoms. Although the structures of MOFs become more complex, the added data greatly assist in the learning. Eventually, we observe a 5% decrease in testing-set MAEs when training is moved from on the QMOF-2 dataset to the QMOF-3 dataset. These results demonstrate the generality and extensibility of the MGANN model.

Fig. 3 Performance of the MGANN model on the QMOF-2 dataset. (a) Comparison of the MAEs and ρ in the bandgaps between the MGANN model and prior benchmark models. All models have been simply divided into two categories: classical machine learning (CML, blue boxes) models with hand-crafted descriptors and deep learning (DL, yellow boxes) models with learnable representations. Note that all results presented here are the average of five parallel runs, with standard deviations illustrated as error bars. (b) Comparison of DFT-computed (from the QMOF-2 database) and MGANN-predicted bandgaps of MOFs on a test set.

Visualization of MGANN latent space. The latent features inside DL models usually serve as the final learned representations of molecules and materials. Gaining insights into the learned latent space is essential for efficient data mining and analysis, which is most often in relation to rational materials design and accelerated materials discovery. We use the unsupervised *t*-distributed stochastic neighbor embedding (*t*-SNE)⁸¹ technique to project the high-dimensional latent features into 2D space for visualizations. As illustrated in Fig. 4, the structure distribution of MOFs in the 2D latent representation space shows a discernible pattern with respect to property values. MOFs with high and low bandgaps are located in the distinctly different regions. It is shown that the learned representations from our MGANN model are leading to groupings of those MOFs that share similarities in their atomic structures, elemental compositions, and chemical properties. We first take the MOF-74-type analogs as an example. The MOF-74 analogs are derived from the known MOF-74 ($\text{Zn}_2(2,5\text{-dihydroxybenzene-1,4-dicarboxylate})$)⁸². They share the same underlying topology but they have difference in metal clusters (e.g., Zn, Mg), functional groups and length of ligands. There are dozens of the MOF-74 analogs⁸³ contained in the QMOF database. The region A (Fig. 4) shows those MOF-74 analogs with $1.6\text{ eV} < E_g < 3\text{ eV}$ are clustering into a group where each structure is highlighted by a white pentagram. Within the grouping region, the different MOF-74 analogs are further discriminated by their bandgaps with generally increasing values from right to left (e.g., $E_{g,3} > E_{g,1} > E_{g,2}$). It demonstrates the capability of the MGANN model to understand similarities in structures and trends in material properties.

Another exemplification is also performed for three isostructural MOFs, i.e., $\text{Cu}[\text{Cu}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ ($\text{pdt}^{2-} = 2,3\text{-pyrazinedithiolate}$)⁸⁴, $\text{Cu}[\text{Ni}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ ⁸⁵, and $\text{Cu}[\text{Ni}(\text{pdt})_2]$ ⁸⁵. The difference between $\text{Cu}[\text{Cu}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ and $\text{Cu}[\text{Ni}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ is in the metal clusters, while the difference between $\text{Cu}[\text{Ni}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ and $\text{Cu}[\text{Ni}(\text{pdt})_2]$ is in the additional C_2H_2 molecules. The representative points of the three MOFs are highlighted by red hollow squares as shown in Fig. 4. It is illustrated that the three representative points generally gather together because of their structural similarity. Interestingly, the various distances between the representative points explicitly indicate the difference of their properties. Specifically, the bandgap values of $\text{Cu}[\text{Cu}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$, $\text{Cu}[\text{Ni}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$, and $\text{Cu}[\text{Ni}(\text{pdt})_2]$ are 0.2395 eV, 0.0392 eV, and 0.0246 eV at the DFT+PBE-D3(BJ) level, respectively⁴⁷. Hence, we can obviously observe that the two

points representing $\text{Cu}[\text{Ni}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ and $\text{Cu}[\text{Ni}(\text{pdt})_2]$ are adjacent due to their close bandgap values, while the $\text{Cu}[\text{Cu}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ point with a larger bandgap is farther apart from the other two. These findings collectively substantiate that the representations generated from the MGANN model are sensitive to the atomic structures, elemental compositions, and chemical properties of materials. This powerful tool enables one to identify promising underlying topologies and infer local chemical trends with respect to various elemental compositions. Eventually, rational materials design can be achieved within the candidate space that are sharply narrowed down.

Fig. 4 *t*-SNE analysis of latent space for MGANN. Dozens of the MOF-74 analogs located in the region A are marked by white pentagrams. The red hollow squares highlight three isostructural MOFs, i.e., $\text{Cu}[\text{Cu}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ (B), $\text{Cu}[\text{Ni}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ (C), and $\text{Cu}[\text{Ni}(\text{pdt})_2]$ (D). Atom colouring: copper = blue, nickel = gray, magnesium = orange, carbon = brown, oxygen = red, hydrogen = pink, nitrogen = stone blue, sulfur = yellow.

CONCLUSIONS

In summary, we proposed a multiscale graph attention neural network architecture for hierarchically learning deep representation of materials, aiming to address the losing long-range correlations between atoms from present local graph representations. The introduction of the self-attention mechanism enables our model to capture multi-scale characteristics of systems over

potentially long distances on top of the local graph. In the meanwhile, this method can keep high parallelizable efficiency during training and predicting. The SOTA performance achieved by our model on predicting the quantum-chemistry properties of MOFs demonstrates its generality and extensibility for complex materials with large unit cells and widely diverse structures and compositions. Moreover, the latent space analysis substantiates the high fidelity of our model in chemical clustering for materials that share chemical and structural similarities. Conclusively, our model makes it possible to explore the latent representations for gaining more chemical insights and useful knowledge to sharply narrow down the search space for high throughput screening, towards accelerating the discovery of complex materials.

DECLARATION OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A., Machine learning for molecular and materials science. *Nature* **2018**, *559* (7715), 547-555.
2. Sanchez-Lengeling, B.; Aspuru-Guzik, A., Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361* (6400), 360.
3. Deringer, V. L.; Caro, M. A.; Csányi, G., Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* **2019**, *31* (46), 1902765.
4. Behler, J.; Parrinello, M., Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401.
5. Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G., Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104* (13), 136403.
6. Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W., Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), 143001.
7. Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G., Gaussian Process Regression for Materials and Molecules. *Chem. Rev.* **2021**, *121* (16), 10073-10141.
8. Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R., Machine Learning Force Fields. *Chem. Rev.* **2021**, *121* (16), 10142-10186.
9. Friederich, P.; Häse, F.; Proppe, J.; Aspuru-Guzik, A., Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **2021**, *20* (6), 750-761.

10. Deringer, V. L.; Bernstein, N.; Csányi, G.; Ben Mahmoud, C.; Ceriotti, M.; Wilson, M.; Drabold, D. A.; Elliott, S. R., Origins of structural and electronic transitions in disordered silicon. *Nature* **2021**, *589* (7840), 59-64.
11. Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A., Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Sci.* **2018**, *4* (2), 268-276.
12. Bernstein, N.; Csányi, G.; Deringer, V. L., De novo exploration and self-guided learning of potential-energy surfaces. *npj Comput. Mater.* **2019**, *5* (1), 99.
13. Zhong, M.; Tran, K.; Min, Y.; Wang, C.; Wang, Z.; Dinh, C.-T.; De Luna, P.; Yu, Z.; Rasouli, A. S.; Brodersen, P.; Sun, S.; Voznyy, O.; Tan, C.-S.; Askerka, M.; Che, F.; Liu, M.; Seifitokaldani, A.; Pang, Y.; Lo, S.-C.; Ip, A.; Ulissi, Z.; Sargent, E. H., Accelerated discovery of CO₂ electrocatalysts using active machine learning. *Nature* **2020**, *581* (7807), 178-183.
14. Chen, A.; Zhang, X.; Zhou, Z., Machine learning: Accelerating materials development for energy storage and conversion. *InfoMat* **2020**, *2* (3), 553-576.
15. Yao, Z.; Sánchez-Lengeling, B.; Bobbitt, N. S.; Bucior, B. J.; Kumar, S. G. H.; Collins, S. P.; Burns, T.; Woo, T. K.; Farha, O. K.; Snurr, R. Q.; Aspuru-Guzik, A., Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nat. Mach. Intell.* **2021**, *3* (1), 76-86.
16. Liu, Y.-B.; Yang, J.-Y.; Xin, G.-M.; Liu, L.-H.; Csányi, G.; Cao, B.-Y., Machine learning interatomic potential developed for molecular simulations on thermal properties of β -Ga₂O₃. *J. Chem. Phys.* **2020**, *153* (14), 144501.
17. Grisafi, A.; Wilkins, D. M.; Csányi, G.; Ceriotti, M., Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems. *Phys. Rev. Lett.* **2018**, *120* (3), 036002.
18. Liu, Y.; Hong, W.; Cao, B., Machine learning for predicting thermodynamic properties of pure fluids and their mixtures. *Energy* **2019**, *188*, 116091.
19. Zafari, M.; Kumar, D.; Umer, M.; Kim, K. S., Machine learning-based high throughput screening for nitrogen fixation on boron-doped single atom catalysts. *J. Mater. Chem. A* **2020**, *8* (10), 5209-5216.
20. Fujii, S.; Yokoi, T.; Fisher, C. A. J.; Moriwake, H.; Yoshiya, M., Quantitative prediction of grain boundary thermal conductivities from local atomic environments. *Nat. Commun.* **2020**, *11* (1), 1854.
21. Behler, J., First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Int. Edit.* **2017**, *56* (42), 12828-12840.
22. Bartók, A. P.; Kondor, R.; Csányi, G., On representing chemical environments. *Phys. Rev. B* **2013**, *87* (18), 184115.
23. Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A., Fast and Accurate Modeling of Molecular Atomization

Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108* (5), 058301.

24. Behler, J., Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134* (7), 074106.
25. Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C., Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **2014**, *89* (9), 094104.
26. LeCun, Y.; Bengio, Y.; Hinton, G., Deep learning. *Nature* **2015**, *521* (7553), 436-444.
27. Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A., Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8* (1), 13890.
28. Smith, J. S.; Isayev, O.; Roitberg, A. E., ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8* (4), 3192-3203.
29. Xie, T.; Grossman, J. C., Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120* (14), 145301.
30. Schütt, K. T.; Sauceda, H. E.; Kindermans, P. J.; Tkatchenko, A.; Müller, K. R., SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148* (24), 241722.
31. Unke, O. T.; Meuwly, M., PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15* (6), 3678-3693.
32. Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O., Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* **2019**, *5* (8), eaav6490.
33. Zubatyuk, R.; Smith, J. S.; Nebgen, B. T.; Tretiak, S.; Isayev, O., Teaching a neural network to attach and detach electrons from molecules. *Nat. Commun.* **2021**, *12* (1), 4870.
34. Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Ong, S. P., Learning properties of ordered and disordered materials from multi-fidelity data. *Nat. Comput. Sci.* **2021**, *1* (1), 46-53.
35. Fung, V.; Zhang, J.; Juarez, E.; Sumpter, B. G., Benchmarking graph neural networks for materials chemistry. *npj Comput. Mater.* **2021**, *7* (1), 84.
36. Wu, Z.; Ramsundar, B.; Feinberg, Evan N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V., MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9* (2), 513-530.
37. Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P., Molecular graph convolutions: moving beyond fingerprints. *J. Comput.-Aided Mol. Des.* **2016**, *30* (8), 595-608.
38. Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P., Graph Networks as a Universal Machine Learning Framework for

Molecules and Crystals. *Chem. Mater.* **2019**, *31* (9), 3564-3572.

39. Prodan, E.; Kohn, W., Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (33), 11635.
40. Grisafi, A.; Nigam, J.; Ceriotti, M., Multi-scale approach for the prediction of atomic scale properties. *Chem. Sci.* **2021**, *12* (6), 2078-2090.
41. Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M., Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **2017**, *3* (12), e1701816.
42. Musil, F.; Grisafi, A.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M., Physics-Inspired Structural Representations for Molecules and Materials. *Chem. Rev.* **2021**, *121* (16), 9759-9815.
43. Hoshino, T.; Papanikolaou, N.; Zeller, R.; Dederichs, P. H.; Asato, M.; Asada, T.; Stefanou, N., First-principles calculations for vacancy formation energies in Cu and Al; non-local effect beyond the LSDA and lattice distortion. *Comp. Mater. Sci.* **1999**, *14* (1), 56-61.
44. Veit, M.; Wilkins, D. M.; Yang, Y.; DiStasio, R. A.; Ceriotti, M., Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *J. Chem. Phys.* **2020**, *153* (2), 024113.
45. Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J., A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12* (1), 398.
46. Behler, J.; Csányi, G., Machine learning potentials for extended systems: a perspective. *Eur. Phys. J. B* **2021**, *94* (7), 142.
47. Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q., Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter* **2021**, *4* (5), 1578-1597.
48. Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R., Carbon Dioxide Capture in Metal–Organic Frameworks. *Chem. Rev.* **2012**, *112* (2), 724-781.
49. Stavila, V.; Talin, A. A.; Allendorf, M. D., MOF-based electronic and opto-electronic devices. *Chem. Soc. Rev.* **2014**, *43* (16), 5994-6010.
50. Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B., Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.* **2020**, *120* (16), 8066-8129.
51. Fan, Y.; Liu, Z.; Chen, G., Recent Progress in Designing Thermoelectric Metal–Organic Frameworks. *Small* **2021**, 2100505.
52. Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D., Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future. *Chem. Mat.* **2017**, *29* (7), 2618-2625.

53. Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M., The Chemistry and Applications of Metal-Organic Frameworks. *Science* **2013**, *341* (6149), 974-+.
54. Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E., Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, 2017; Vol. 70, pp 1263--1272.
55. Park, C. W.; Kornbluth, M.; Vandermause, J.; Wolverton, C.; Kozinsky, B.; Mailoa, J. P., Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *npj Comput. Mater.* **2021**, *7* (1), 73.
56. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; Solomon, J. M., Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans. Graph.* **2019**, *38* (5), 146.
57. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I., Attention is all you need. In *Advances in neural information processing systems*, 2017; pp 5998-6008.
58. Lee, A. A.; Yang, Q.; Sresht, V.; Bolgar, P.; Hou, X.; Klug-McLeod, J. L.; Butler, C. R., Molecular Transformer unifies reaction prediction and retrosynthesis across pharma chemical space. *Chem. Commun.* **2019**, *55* (81), 12152-12155.
59. Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A., Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Central Sci.* **2019**, *5* (9), 1572-1583.
60. Pesciullesi, G.; Schwaller, P.; Laino, T.; Reymond, J.-L., Transfer learning enables the molecular transformer to predict regio- and stereoselective reactions on carbohydrates. *Nat. Commun.* **2020**, *11* (1), 4874.
61. Schwaller, P.; Probst, D.; Vaucher, A. C.; Nair, V. H.; Kreutter, D.; Laino, T.; Reymond, J.-L., Mapping the space of chemical reactions using attention-based neural networks. *Nat. Mach. Intell.* **2021**, *3* (2), 144-152.
62. Weininger, D., SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31-36.
63. Weininger, D.; Weininger, A.; Weininger, J. L., SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29* (2), 97-101.
64. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y., Spectral networks and deep locally connected networks on graphs. In *2nd International Conference on Learning Representations*, 2014.
65. Guo, M.-H.; Cai, J.-X.; Liu, Z.-N.; Mu, T.-J.; Martin, R. R.; Hu, S.-M., PCT: Point cloud transformer. *Comp. Visual Media* **2021**, *7* (2), 187-199.
66. Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M., Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science* **2002**, *295* (5554), 469-472.
67. Deng, H. X.; Grunder, S.; Cordova, K. E.; Valente, C.; Furukawa, H.; Hmadeh, M.; Gandara, F.; Whalley, A. C.; Liu, Z.; Asahina, S.; Kazumori, H.; O'Keeffe, M.; Terasaki, O.; Stoddart, J. F.; Yaghi, O. M., Large-Pore Apertures

in a Series of Metal-Organic Frameworks. *Science* **2012**, 336 (6084), 1018-1023.

68. Kalmutzki, M. J.; Hanikel, N.; Yaghi, O. M., Secondary building units as the turning point in the development of the reticular chemistry of MOFs. *Sci. Adv.* **2018**, 4 (10).
69. Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q., Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, 64 (12), 5985-5998.
70. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, 77 (18), 3865-3868.
71. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, 132 (15), 154104.
72. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, 32 (7), 1456-1465.
73. Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Ruhl, S.; Wolverton, C., The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **2015**, 1.
74. Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q., Computation-Ready, Experimental Metal–Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mat.* **2014**, 26 (21), 6185-6192.
75. Sun, L.; Campbell, M. G.; Dincă, M., Electrically Conductive Porous Metal–Organic Frameworks. *Angew. Chem. Int. Edit.* **2016**, 55 (11), 3566-3579.
76. Xie, L. S.; Sun, L.; Wan, R.; Park, S. S.; DeGayner, J. A.; Hendon, C. H.; Dincă, M., Tunable Mixed-Valence Doping toward Record Electrical Conductivity in a Three-Dimensional Metal–Organic Framework. *J. Am. Chem. Soc.* **2018**, 140 (24), 7411-7414.
77. Faber, F.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R., Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem.* **2015**, 115 (16), 1094-1101.
78. He, Y.; Cubuk, E. D.; Allendorf, M. D.; Reed, E. J., Metallic Metal–Organic Frameworks Predicted by the Combination of Machine Learning Methods and Ab Initio Calculations. *J. Phys. Chem. Lett.* **2018**, 9 (16), 4562-4569.
79. Lam Pham, T.; Kino, H.; Terakura, K.; Miyake, T.; Tsuda, K.; Takigawa, I.; Chi Dam, H., Machine learning reveals orbital interaction in materials. *Sci. Technol. Adv. Mater.* **2017**, 18 (1), 756-765.
80. Lysogorskiy, Y.; Oord, C. v. d.; Bochkarev, A.; Menon, S.; Rinaldi, M.; Hammerschmidt, T.; Mrovec, M.;

- Thompson, A.; Csányi, G.; Ortner, C.; Drautz, R., Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon. *npj Comput. Mater.* **2021**, 7 (1), 97.
81. van der Maaten, L.; Hinton, G., Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, 9, 2579-2605.
82. Rosi, N. L.; Kim, J.; Eddaoudi, M.; Chen, B.; O'Keeffe, M.; Yaghi, O. M., Rod Packings and Metal–Organic Frameworks Constructed from Rod-Shaped Secondary Building Units. *J. Am. Chem. Soc.* **2005**, 127 (5), 1504-1518.
83. Witman, M.; Ling, S.; Anderson, S.; Tong, L.; Stylianou, K. C.; Slater, B.; Smit, B.; Haranczyk, M., In silico design and screening of hypothetical MOF-74 analogs and their experimental synthesis. *Chem. Sci.* **2016**, 7 (9), 6263-6272.
84. Peng, Y.-L.; Pham, T.; Li, P.; Wang, T.; Chen, Y.; Chen, K.-J.; Forrest, K. A.; Space, B.; Cheng, P.; Zaworotko, M. J.; Zhang, Z., Robust Ultramicroporous Metal–Organic Frameworks with Benchmark Affinity for Acetylene. *Angew. Chem. Int. Ed.* **2018**, 57 (34), 10971-10975.
85. Aubrey, M. L.; Kapelewski, M. T.; Melville, J. F.; Oktawiec, J.; Presti, D.; Gagliardi, L.; Long, J. R., Chemiresistive Detection of Gaseous Hydrocarbons and Interrogation of Charge Transport in Cu[Ni(2,3-pyrazinedithiolate)₂] by Gas Adsorption. *J. Am. Chem. Soc.* **2019**, 141 (12), 5005-5013.