

Structure motif-centric learning framework for inorganic crystalline systems

*Huta R. Banjade^{1†}, Sandro Hauri^{2†}, Shanshan Zhang², Francesco Ricci³, Weiyi Gong¹,
Geoffroy Hautier^{3,4}, Slobodan Vucetic^{2*}, Qimin Yan^{1*}*

¹Department of Physics, Temple University, Philadelphia, PA 19122, USA. ²Department of Computer and Information Science, Temple University, Philadelphia, PA 19122, USA. ³Institute of Condensed Matter and Nanoscience (IMCN), Université catholique de Louvain (UCLouvain), Chemin étoiles 8, bte L7.03.01, Louvain-la-Neuve 1348, Belgium. ⁴Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA.

*†*These authors contributed equally to this work.

Incorporation of physical principles in a machine learning (ML) architecture is a fundamental step toward the continued development of artificial intelligence for inorganic materials. As inspired by the Pauling's rule, we propose that structure motifs in inorganic crystals can serve as a central input to a machine learning framework. We demonstrated that the presence of structure motifs and their connections in a large set of crystalline compounds can be converted into unique vector representations using an unsupervised learning algorithm. To demonstrate the use of structure motif information, a motif-centric learning framework is created by combining motif information with the atom-based graph neural networks to form an atom-motif dual graph network (AMDNet), which is more accurate in predicting the electronic structures of metal oxides such as bandgaps. The work illustrates the route toward fundamental design of graph neural network learning architecture for complex materials by incorporating beyond-atom physical principles.