Self-supervised GNN for clustering of two-dimensional materials

Ivan Savochkin[a], **Yihao Wei**[a], **Anupam Bhattacharya**[a], **Artem Mishchenko**[a]

[a] *Department of Physics and Astronomy, University of Manchester, United Kingdom M13 9PL* yihao.wei@postgrad.manchester.ac.uk, anupam.bhattacharya@manchester.ac.uk, artem.mishchenko@manchester.ac.uk

Two-dimensional (2D) materials have garnered significant attention due to their exceptional electronic, optical, and mechanical properties [1, 2]. For example, graphene's well-defined hexagonal lattice yields extraordinary electron mobility and mechanical strength [3, 4], while transition metal dichalcogenides (TMDs), such as $NbSe_2$, exhibit distinct behaviors like superconductivity, charge density wave and spin-orbit coupling that induce emergence of topological electronics and strong correlated phenomena [5, 6, 7]. Although many 2D materials appear structurally simple, their subtle differences—such as variations in crystal symmetry, stacking order, and defect structures—necessitate precise characterization methods. Effective encoding of structural information is of much significance in developing machine learning approaches in the study of 2D materials.

In this study, we present a self-supervised framework that leverages message passing graph neural networks (GNNs) to extract robust structural fingerprints from two-dimensional (2D) materials [8, 9]. The pipeline of our method is shown in Figure 1. Our method models each material as a graph, where atoms are represented as nodes characterized by relative Cartesian coordinates, and edges encode both nearest and next-nearest neighbor interactions, including periodic boundary conditions via supercell construction (e.g., 3×3 or 4×4 grids). Additionally, angular information is incorporated through hyperedges that connect triplets of atoms, thereby capturing essential geometric features of the crystal structure [10].

We use random augmentation achieved by applying controlled random perturbations to atomic positions and varying the supercell dimensions to ensure that the learned representations remain invariant to minor structural fluctuations while preserving critical material properties. The core of our approach is a contrastive learning scheme based on the InfoNCE loss function [11], which drives the model to minimize the distance between embeddings of perturbed versions of the same structure while maximizing the separation between embeddings of distinct structures. This strategy effectively mitigates trivial solutions and enhances the discriminative power of the resulting fingerprints.

Post-training, the extracted embeddings are subjected to clustering analyses using techniques such as t-SNE and HDBSCAN. The clusters that emerge correlate strongly with key structural parameters, including lattice symmetry, unit cell composition, chemical elements and interatomic distances. Our results indicate that the proposed self-supervised approach not only captures intrinsic geometric and chemical features but also provides a scalable pathway for high-throughput material screening and property prediction.

Our method as a distinct alternative to traditional supervised approaches such as those in CGCNN [8] and ALIGNN [12] provides a notably streamlined model that significantly reduces computational overhead. Our model is around 2 orders of magnitude quicker than CGCNN and nearly 4 orders of magnitude faster than ALIGNN. This efficiency gain demonstrates the potential of our approach for scalable, high-throughput applications in materials science.

References

[1] K S Novoselov, A Mishchenko, A Carvalho, and A H Castro Neto. 2D materials and van der waals heterostructures. *Science*, 353(6298):aac9439, July 2016.

[2] Brian C Wyatt, Andreas Rosenkranz, and Babak Anasori. 2D MXenes: Tunable mechanical and tribological properties. *Adv. Mater.*, 33(17):e2007973, April 2021.

[3] Antonio H Castro Neto, Francisco Guinea, Nuno MR Peres, Kostya S Novoselov, and Andre K Geim. The electronic properties of graphene. *Reviews of modern physics*, 81(1):109–162, 2009.

[4] Deji Akinwande, Christopher J Brennan, J Scott Bunch, Philip Egberts, Jonathan R Felts, Huajian Gao, Rui Huang, Joon-Seok Kim, Teng Li, Yao Li, et al. A review on mechanics and mechanical properties of 2d materials—graphene and beyond. *Extreme Mechanics Letters*, 13:42–77, 2017.

[5] Ying Xing, Kun Zhao, Pujia Shan, Feipeng Zheng, Yangwei Zhang, Hailong Fu, Yi Liu, Mingliang Tian, Chuanying Xi, Haiwen Liu, et al. Ising superconductivity and quantum phase transition in macro-size monolayer nbse2. *Nano letters*, 17(11):6802–6807, 2017.

[6] Xiaoxiang Xi, Liang Zhao, Zefang Wang, Helmuth Berger, László Forró, Jie Shan, and Kin Fai Mak. Strongly enhanced charge-density-wave

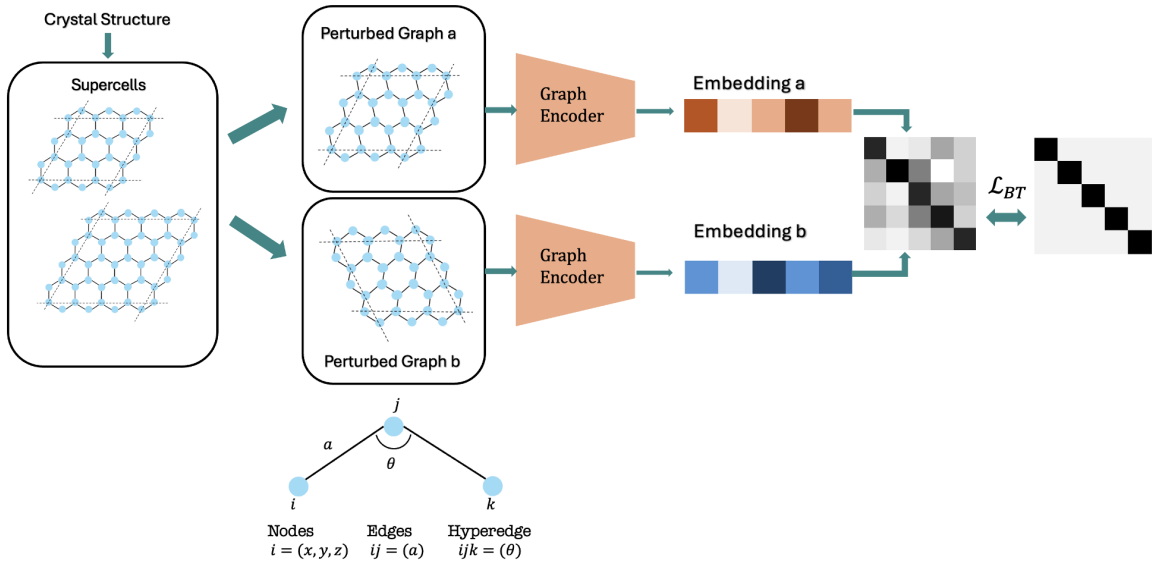

Fig. 1: Overall pipeline of the self-supervised framework.

order in monolayer nbse2. *Nature nanotechnology*, 10(9):765–769, 2015.

[7] Yuki Nakata, Katsuaki Sugawara, Ryota Shimizu, Yoshinori Okada, Patrick Han, Taro Hitosugi, Keiji Ueno, Takafumi Sato, and Takashi Takahashi. Monolayer 1t-nbse2 as a mott insulator. *NPG Asia Materials*, 8(11):e321–e321, 2016.

[8] Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.*, 120:145301, Apr 2018.

[9] Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, et al. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93, 2022.

[10] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3558–3565, 2019.

[11] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding, 2019.

[12] Kamal Choudhary and Brian DeCost. Atomistic line graph neural network for improved materials property predictions. *Npj Comput. Mater.*, 7(1), November 2021.