

GRAPH NEURAL NETWORK BASED SCREENING OF METAL-ORGANIC FRAMEWORKS FOR CO₂ CAPTURE

Zikri Bayraktar, Mengying Li & Shahnawaz Molla

Schlumberger-Doll Research

Cambridge, MA 02139, USA

{zbayraktar,mli50,smolla}@slb.com

ABSTRACT

Our ability to capture and remove carbon dioxide (CO₂) at gigaton scale within a decade or two depends on our ability to quickly identify new materials that are high performing, selective over other gases with low energy demand and then further develop them for large scale deployment. As a proven technology for gas separation in other industrial applications, metal-organic frameworks (MOF) come in virtually unlimited number of crystal combinations in their highly porous lattice and may offer the solution for CO₂ capture from atmosphere or industrial point sources. Although MOFs can have highly complex crystal structure, which cannot be easily exploited in tabular data format in conventional ML methods or more recent Deep Learning methods, Graph Neural Networks can easily be trained on their representative crystallographic information file (CIF) content. In this work, we train GNNs to create an end-to-end workflow to screen large number of MOF crystal structures directly from the data within the crystallographic information files for their CO₂ working capacity or CO₂/N₂ selectivity under low-pressure conditions. Our preliminary results show that a simple 2-layered Graph Convolution Networks (GCN) can easily achieve R² score in the range of 0.87 to 0.89, easily.

1 NEED FOR CARBON DIOXIDE CAPTURE

Carbon dioxide (CO₂) capture, utilization, and sequestration (CCUS) is expected to play an important role in the global campaign to limit global warming to 1.5 °C. However, to achieve this ambitious goal, we need to deploy CO₂ capture capabilities at gigaton capacity - approximately 150X increment by 2050 (Global CCS Institute, 2020).

Currently, CO₂ capture using chemical solvents is the most mature technology that has been proven to work at large-scale (Leung et al., 2014). However, these capture systems have large footprints and the energy demand to operate the process is very high. Studies conducted by U.S. National Energy Technology Laboratory (NETL) suggests that implementing aqueous amine-based CO₂ capture can reduce the energy efficiency of a coal-fired power plant by 20-30% (Oko et al., 2017). The use of large amount of thermal energy for solvent regeneration accounts for 20-25% of the operating cost despite deploying advanced solvents.

Solvent-based capture systems are also expensive to build due to the large footprint of the packed bed towers used for CO₂ absorption and desorption. As a result, new separation technologies using either solid sorbents, membranes, or cryogenics are being developed (Abanades et al., 2015) to address the need for new material development to improve the CO₂ capture capacity, selectivity over other gases, and to reduce the regeneration energy demand. Hence, there is urgency to develop energy-efficient and compact capture processes to allow widespread adoption of CCUS in the industrial sector and reduce global warming.

2 METAL-ORGANIC FRAMEWORKS (MOF) FOR CO₂ CAPTURE

A variety of solid sorbent-based materials are being investigated for their CO₂ adsorption capacity, CO₂ selectivity, stability, and regeneration capacity to efficiently remove CO₂ from point sources

(Dziejarski et al., 2023). Specifically, metal-organic frameworks (MOFs), a class of highly porous crystalline materials, have come across as promising candidates due to their exceedingly high CO₂ adsorption capacities that are inherently tunable through their unique structural features (Sumida et al., 2012) (Li et al., 2022) and more importantly, their scalability for industrial applications. MOFs are built through transition metal clusters and organic linkers into structural building units (SBUs) that are further extended into well-defined, uniformed structures with high surface area and structural stability. Those highly tunable SBUs in MOFs offer exceptional adaptability in isorecticular manipulation as well as their controllable functionalities in CO₂ adsorption through their theoretical, infinite combinations among metal ions, organic linkers, functional group, and optimization of pores spacing (Lu et al., 2022)

For post-combustion CO₂ capture processes, there is an urgent need to understand the role of impurities, water in particular, the variability in the concentrations of impurities in the feed stream and more importantly, their impact on the stability of the sorbents as well as the energy requirement for the regeneration processes. However, to-date, only a small number of MOFs have been studied experimentally to meet target-specific requirements, which warrants the necessity for a faster and reliable way to design and screen MOFs for post-combustion capture processes.

Past ML work on MOFs screening was driven by the subject matter experts (SME) time-consuming analysis of the properties through feature engineering with geometric, chemical, and energy-based descriptors. Graph Neural Network, on the other hand, can be used to represent complex crystal structures of MOF using their crystallographic information, including but not limited to their unit cell values, space groups, identities, and position of the atoms, as well as properties of each bond.

3 PROPOSED GRAPH NEURAL NETWORKS (GNN) METHOD

Various types of data in many aspects of our lives can be represented by graphs and there are solutions already implemented using Graph Neural Networks Scarselli et al. (2009). Recommender systems in online shopping sites (Wu et al., 2022), traffic prediction in mobile phone applications (Derrow-Pinion et al., 2021), weather forecasting (Lam et al., 2023) and our beloved social-networks, all use different types of GNNs in their core. GNNs are also utilized in science for physics simulations (Sanchez-Gonzalez et al., 2020), molecule classification (Sanchez-Lengeling et al., 2019) or molecule design (Liu et al., 2018), drug discovery (You et al., 2018), protein discovery (Knutson et al., 2022) and many other tasks.

Our proposed method utilizes the CIF file content and encode it in graph representation to train GNN models to predict material performance as illustrated in Figure 1. Material performances were computed through Grand Canonical Monte Carlo (GCMC) method (Lu et al., 2022). Unlike simple molecules, a single MOF crystal can make up to a thousand of bonds, which brings a challenge for any ML method to handle such nonuniform dataset. In our preliminary results, we have demonstrated GCNs coded in PyG can easily handle the training of such complex data structures as undirected graphs in a regression problem of predicting MOF properties such as the CO₂ working capacity or selectivity. As expected from the problem definition, i.e., screening large number of potential candidates, most of the MOF crystals are low performing. This can be seen by the label distribution for working capacity and selectivity in Figure 2. This creates a challenge as the predicted data is skewed for this regression task.

4 PRELIMINARY RESULTS ON GNN-BASED SCREENING OF MOFS

In this work, we implemented and trained end-to-end GNN models directly from the existing CIF content via PyG and used the corresponding GCMC data to predict their CO₂ working capacity or CO₂/N₂ selectivity as depicted in Figure 1 (Fey et al., 2019). We utilized the CIF files for 340,000 MOFs where over 250,000 of the files were missing their bond distances. Using the unit cell dimensions and atom site fractional coordinates, we calculated the bond distances to use them as one of the node features. Along with the bond distances, one-hot encoded bond types constitute the edge attributes for each MOF. Adjacency matrix which represents which atoms are connected through bonds (i.e., graph edges) was encoded as edge lists. Node attributes include the atomic weight of each atom and one-hot encoded atom types. Crystal unit cell dimensions and angles are also utilized along with the one-hot encoded crystal type information as shown on the left in Figure 3. Exploiting

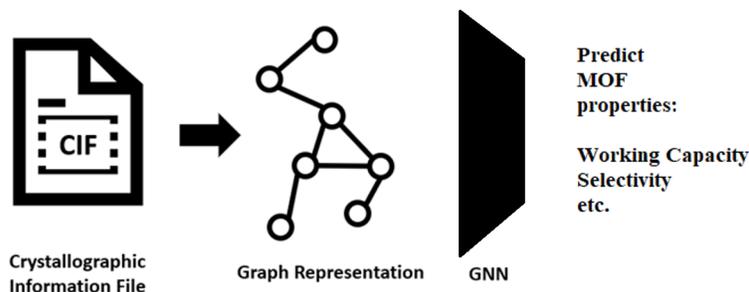


Figure 1: Proposed end-to-end GNN based MOF screening workflow utilizing CIFs to either predict working capacity or selectivity computed by GCMC simulations.

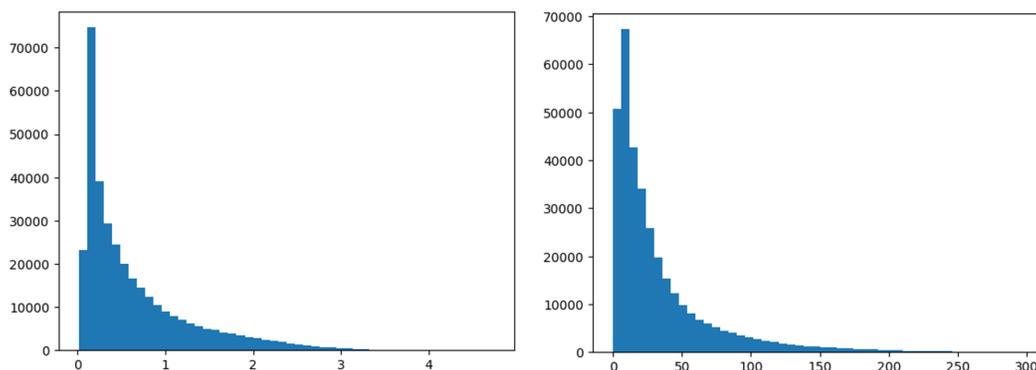


Figure 2: Histogram of the label distribution for working capacity (left) and selectivity (right) for all MOF simulations. Most MOF candidate data is skewed towards low performance hence the need for AI-based high-throughput screening.

the information provided in each CIF files, one can eliminate the need for hand-picked geometric or chemical descriptors and reduce the SMEs workload. Initially, we focused on training GCN models (Coley et al., 2018) and achieved R^2 score ranging 0.87 to 0.89, easily. A sample test result is shown on the right side of Figure 3. In comparison to our preliminary results, prior ML work using only SME-designed geometric descriptors alone reached R^2 score up to 0.75 (Burner, 2020), whereas custom designed atomic property-weighted radial distribution function (AP-RDF) features can push the ML model R^2 score to a range of 0.83 and 0.94, respectively. Similarly, it was shown that CNN based end-to-end screening can achieve R^2 score up to 0.91 (Lu et al., 2022).

In our preliminary results shown above, we explored Graph Convolution Networks (Raschka et al., 2022) (Labonne et al., 2023) and achieved good R^2 score results with simple two-layered convolution architecture utilizing the information coming directly from the CIFs. In the follow-up work, we aim to explore graph attention networks (Velickovic et al., 2017) as comparative study to identify most efficient GNN architectures for the screening task. In addition, implementation of a custom cost functions with weighting based on the density for the regression may help with the performance (Steininger et al., 2021) beyond simple MSE cost function that we have used for GCN training.

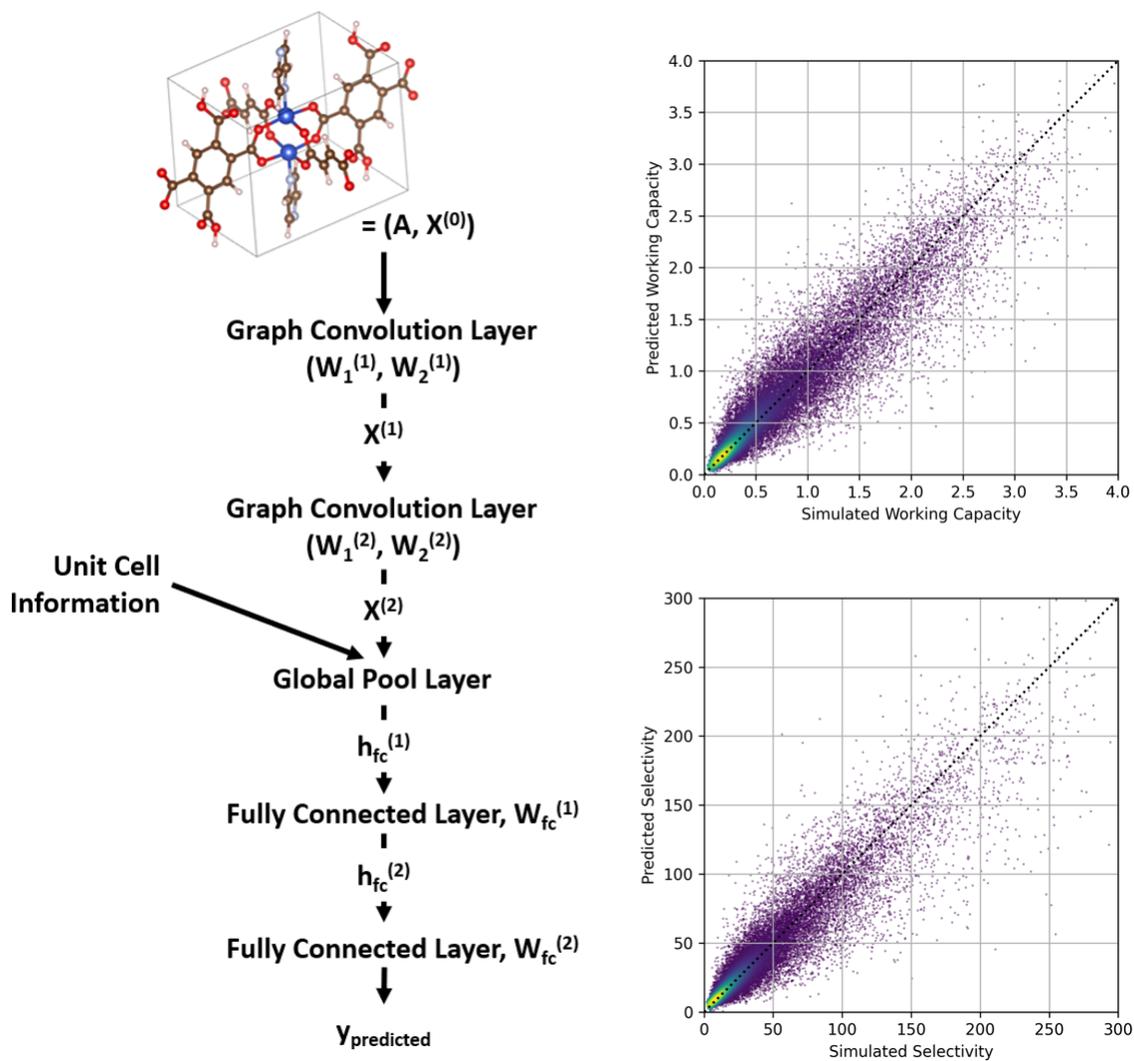


Figure 3: Architecture of the GCN trained on the CIF data (left). Early results for GCN trained for working capacity (top right) and selectivity (bottom right).

REFERENCES

- Abanades, J.C., Arias, B., Lyngfelt, A., Mattisson T., Wiley, D.E., Li, H., Ho, M.T., Mangano, E. & Brandani, S. (2015) Emerging CO₂ capture systems. *International Journal of Greenhouse Gas Control*, 40, 126-166. <https://doi.org/10.1016/j.ijggc.2015.04.018>
- Burner, J., et. al. (2020) High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO₂ Adsorption Properties of MetalOrganic Frameworks. *J. Phys. Chem. C*, 2020, 124, 51, 2799628005.
- Coley, C.W., Jin, W., Rogers, L., Jamison, T.F., Jaakkola, T., Green, W.H., Barzilay, R., & Jensen, K.F.. (2018) A graph-convolutional neural network model for the prediction of chemical reactivity. *Chemical Science*, 10, 370-377, <https://doi.org/10.1039/c8sc04228d>.
- Derrow-Pinion, A., She, J., Wong, D., Lange, O., Hester, T., Perez, L., Nunkesser, M., Seongjae, L., Xueying, G., Wiltshire, B., Battaglia, P.W., Gupta, V., Ang, L., Zhongwen, X., Sanchez-Gonzalez, A., Yujia, L. & Velickovic, P. (2021) ETA Prediction with Graph Neural Networks in Google Maps. *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21)*. Association for Computing Machinery, New York, NY, USA, 37673776, 2021.
- Dziejarski, P., Serafin, J., Anderson, K., Krzyzyska, R., (2023) CO₂ capture materials: a review of current trends and future challenges. *Materials Today Sustainability*, 24, 100483.
- Fey, Matthias, & Lenssen, Jan, E. (2019) Fast Graph Representation Learning with PyTorch Geometric. *2019 ICLR Worksop on Representation Learning on Graphs and Manifolds*.
- Global CCS Institute. (2020) Global Status of CCS Report. *Global Status of CCS Report: 2020*, <https://www.globalccsinstitute.com/resources/publications-reports-research/global-status-of-ccs-report-2020/>
- Knutson, C., Bontha, M., Bilbrey, J.A. et al. Decoding the proteinligand interactions using parallel graph neural networks. *Sci Rep* 12, 7624. <https://doi.org/10.1038/s41598-022-10418-2>.
- Labonne, M. (2023) Hands-on Graph Neural Networks Using Python. *Packt Publishing*, 1st Edition.
- Lam, R., et al., (2023) Learning skillful medium-range global weather forecasting. *Science*, Vol 382, Issue 6677 pp. 1416-1421 DOI: 10.1126/science.adi2336.
- Leung, D.Y.C., Caramanna, G. & Maroto-Valer, M.M.. (2014) An overview of current status of carbon dioxide capture and storage technologies. *Renewable and Sustainable Energy Reviews*, 39, 426-443. <https://doi.org/10.1016/j.rser.2014.07.093>.
- Li, M., Hall, J. N., Fleming, K., and Bollini, P. (2022). Tuning CO₂ Binding in Metal-Organic Framework Materials by Varying Metal Identity and Oxidation State. *Energy & Fuels* 2022, 36 (11), 5816-5824. DOI: 10.1021/acs.energyfuels.2c00443.
- Liu, Q., Allamanis, M., Brockschmidt, M. & Gaunt, A.L. (2018) Constrained graph variational autoencoders for molecule design. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Curran Associates Inc., Red Hook, NY, USA, 78067815.
- Lu, C., Wan, X., Ma, X., Guan, X., & Zhu, A.. (2022) Deep-Learning-Based End-to-End Predictions of CO₂ Capture in Metal-Organic Frameworks. *Journal of Chemical Information and Modeling*, 62 (14), 3281-3290.
- Oko, E., Wang, M. & Joel, A.S.. (2017) Current status and future development of solvent-based carbon capture. *International Journal of Coal Science and Technology*, 4 (1), 5-14. <https://doi.org/10.1007/s40789-017-0159>.
- Raschka, S., Liu, Y., Mirjalili, V., (2022) Machine Learning with PyTorch and Scikit-Learn. *Packt Publishing*, 1st Edition.
- Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J. & Battaglia, P.W.. (2020) Learning to simulate complex physics with graph networks. *Proceedings of the 37th International Conference on Machine Learning, JMLR.org*, Article 784, 84598468.

- Sanchez-Lengeling, B. et al. (2019) Machine Learning for Scent: Learning Generalizable Perceptual Representations of Small Molecules. *arXiv*.
- Scarselli, F., et al. (2009) The graph neural network model. *IEEE Transactions on Neural Networks*, Vol 20(1), pp. 61–80.
- Sumida, K., Rogow, D., Mason, J. A., McDonald, T. M., Block, E. D., Herm, Z. R. Bae, T., Long, J., (2012) Carbon Dioxide Capture in Metal-Organic Frameworks. *Chemical Reviews*, 112(2), 724-781.
- Steininger, M., Kobs, K., Davidson, P. et al. (2021) Density-based weighting for imbalanced regression. *Mach Learn* 110, 21872211. <https://doi.org/10.1007/s10994-021-06023-5>.
- Velickovic, P., et. al. (2017) Graph Attention Networks. *ArXiv*.
- Wu, S., et. al. (2022) Graph Neural Networks in Recommender Systems: A Survey. *ACM Comput. Surv.*37, 4, Article 111.
- You, J., Liu, B., Ying, R., Pande, V., & Leskovec. (2018). Graph convolutional policy network for goal-directed molecular graph generation. *In Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Curran Associates Inc., Red Hook, NY, USA, 64126422.