

Efficient Integration of Molecular Representation and Message-Passing Neural Networks for Predicting Small Molecule Drug-like Properties

Shreyas Bhat Brahmavar^{1,2§}, Mrunmay Mohan Shelar^{1,3§}, Revanth Harinarthini³, Bandaru

Hemanth Sai Krishna³, Nahush Harihar Kumta³, Ojas Wadhvani³, Raviprasad Aduri^{1,4*}

¹Department of Biological Sciences, ²Department of Electrical and Electronics Engineering,

³Department of CS & IS, ⁴APPCAIR, BITS Pilani, K K Birla Goa campus, Goa, India, 403726.

**Keywords --- MPNN, Bond Order Matrix, Semi-Master Node, Drug-likeness, Property
Prediction**

* To whom correspondence should be addressed: aduri@goa.bits-pilani.ac.in

§ These authors share equal contribution and are co-first authors

ABSTRACT

The physicochemical properties of a drug molecule determine its metabolism properties. There have been hybrid quantum mechanics approaches with computer-aided drug design and recent supervised machine-learning approaches to predict these properties of small-molecule drugs. However, these methods are low in accuracy and computationally expensive. To get around this problem and improve the performance of a model that predicts the properties of drug molecules, we came up with a novel architecture that uses a "bond order matrix" and structural information to improve molecular graph representations and information in the molecule. Message-passing neural networks (MPNNs) are a framework used to learn local and global features from irregularly formed data invariant to permutations. We take advantage of MPNN architecture and introduce a "semi-master node," a unique way of representing the functional groups in a small molecule and aggregating features obtained from the functional groups, in anticipation of reverse engineering small molecules given the desired physicochemical properties. This novel architecture and molecule representation were evaluated on the QM9 dataset, which has 133,000 stable small organic molecules with nine heavy atoms (CONF) out of the GDB-17 chemical universe. The metric for evaluating the model's performance is DFT error, an estimated average error of the properties of each molecule. Our models have shown a performance gain of ~10%.

1. INTRODUCTION

“Sequence dictates structure dictates function” is an axiom applicable to both biological macromolecules and small molecules. Most drugs are made of small molecules that target the biological macromolecule proteins. Though Structure-Based Drug Design (SBDD) has significantly improved the traditional methods of drug development, the attrition rates in SBDD are still remarkably high and often result in financial and time losses. Two major players in the drug design are the biological target itself and the small molecule. It is noteworthy that the biological target space is a lot more limited compared to the small molecule chemical space. Hence, it is not surprising that the physicochemical properties of the small molecule are often at the forefront of the small molecule being ‘druglike’. Lipinski’s rule of five indicates the ‘drug-likeness’ of small molecules¹. However, several known FDA-approved drugs do not conform to these rules². Moreover, following these rules also limits the chemical space that can be explored for developing novel drug candidates³.

The binding affinity of a small molecule to a given biological target and its Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties are all intrinsic to its structure. The physicochemical properties of a given small molecule, such as potential energy, dipole moment, atomization energy, enthalpy of dissociation, and HOMO/LUMO energies, dictate these various ‘drug-likeness’ properties. Experimental analysis, both in vitro and in vivo, of these properties is time-consuming and extremely expensive. However, most of these properties can be calculated from the sequence of the small molecule using quantum mechanical (QM) methods. However, Most of these techniques rely on approximations because of the

complexity of using quantum mechanical principles for calculations. A few notable techniques include DFT (Density functional theory) ⁴, the GW approximation ⁵, and Quantum Monte-Carlo ⁶. These techniques required computation with a runtime of $O(N^3)$ (where N is the number of electrons), which can explode and go as high as eight hours for a 17-heavy atom molecule (when run on a single core of a Xeon E5-2660 (2.2 GHz) using a version of Gaussian G09 (ES64L-G09RevD.01) (Bing et al., 2017). Additionally, the techniques weren't entirely accurate either, further contributing to the frustration of a large runtime.

The traditional process of finding out the properties of small molecules using quantum mechanical computation approaches can be computationally expensive. Owing to the limitations of traditional methods mentioned above, it made sense to use Message Passing Neural Networks (NN) to capture structural information to predict these properties. There is ample data and enough computational resources to train an NN model today. Thus, it does not come as a surprise that individuals would try to shift their approach in the direction of using Neural Networks. Two prominent approaches by Behler *et al.* ⁷ and Rupp *et al.* ⁸. Rupp *et al.*, have attempted to estimate the solution to quantum mechanics using Neural Networks (NN) without directly appealing to DFT ⁸. Both these approaches used hand-engineered features (e.g.: symmetry functions, Coulomb matrix) that ensured physical aspects were built into their input representations.

Several methods have been successful in using neural network models to learn molecule symmetries. Some prominent examples used preprocessing steps to extract canonical graph features/symmetries, which are then fed into classifiers (ML models) for training, such as in Niepert et al. ⁹ and Rupp et al.⁸. However, Gilmer et al. ¹⁰ have proposed an all-encompassing,

common framework called Message Passing Neural Network (MPNN), where the model learns a message-passing algorithm followed by an aggregation procedure and a readout phase. They have benchmarked this approach on the QM9 dataset. Another similar example includes that of Scarselli et al.,¹¹ where they defined a message passing process on graphs that runs until train and validation losses converge and not for a finite number of steps as in the MPNN. The two major components of these methods are the way the data is presented to the machine, which is generally represented as fingerprints, string representations, atom distance maps, but neither contains enough relevant information to specify chemical structure fully, and the second is the architecture of the ML/DL itself.

To address challenges in the representation of molecules that affect the efficiency of the model, in this paper, we present a novel MPNN architecture using a ‘bond order matrix’ (Fig 1) representation of small molecules. In this context, we provide the following key contributions:

- We developed a more robust input for the ML models and a new bond order matrix (Fig 1) representation for the small molecules such that no information related to the nature and type of bonding is lost. Moreover, we also tried to experiment with whether there is a need to generate a matrix and SMILES notation by itself as input is sufficient.
- Added a semi-master node (Fig 2) to the MPNN architecture that creates aromatic features, atom features, and bond features as an input to the Message Passing Neural Network. Herein a semi-master node (Fig 2) was added to input the structural information such as rings, aromaticity, etc. This addition proved vital since structure affects many properties like dipole moment, potential energy, etc.

- This method of enhancing the representation of molecules as input to the MPNN yields an efficient feature representation, saving compute over time and memory as compared to previous MPNNs used for molecule representation for this task of property predictions. (Figure 3).

Here we focus on the QM9 dataset, which consists of 130k molecules with 13 different physicochemical properties computed using DFT calculations. The metric we used to measure the performance of our model is called “DFT error”, which is an estimated average error of the properties of each molecule.

2. MESSAGE PASSING NEURAL NETWORKS

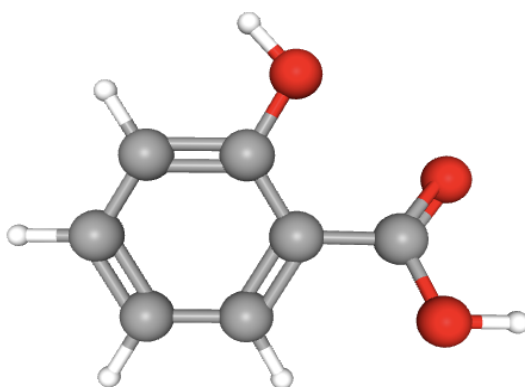
MPNNs operate on undirected graphs G with node features x_v and edge features e_{vw} . The forward pass has two phases: the message passing phase and the readout phase. The message passing phase runs for T time steps and is defined in terms of message function M_t and vertex update functions U_t . Then, finally, the readout phase computes a feature vector for the whole graph using the readout function R .

2.1 BOND ORDER (BO) MATRIX: INPUT REPRESENTATION

This new matrix to be generated should preserve the properties of the adjacency matrix (distance matrix can be calculated from the adjacency matrix) and address the kind of bond between a pair of atoms.

The bond between atoms:

The bond between a pair of atoms can be marked using bond order. Bond order in organic chemistry is the number of covalent bonds shared between a pair of atoms. In the new square matrix, an element denotes the bond order between the atoms by simply encoding the bonds as 3 for a triple bond, 2 for a double bond, 1 for a single bond, and 0.5 for a dative bond. If an element in the BO matrix is more significant than zero, it implies that the pair of atoms are adjacently preserving the property of the adjacency matrix. The bond order property of an element in the bond order matrix (Figure 1) marks the kind of bond shared between the atoms.



	C	C	C	C	C	C	C	O	O	O
C	0	1.5	0	0	0	1.5	0	0	0	0
C	1.5	0	1.5	0	0	0	0	0	0	0
C	0	1.5	0	1.5	0	0	0	0	0	0
C	0	0	1.5	0	1.5	0	0	0	0	1
C	0	0	0	1.5	0	1.5	1	0	0	0
C	1.5	0	0	0	1.5	0	0	0	0	0
C	0	0	0	0	1	0	0	2	1	0
O	0	0	0	0	0	0	2	0	0	0
O	0	0	0	0	0	0	1	0	0	0
O	0	0	0	1	0	0	0	0	0	0

Figure 1. An example of expressing a molecule (Salicylic acid) in its Bond-Order representation.

2.2 Message and Update Stage

Each edge is projected to a larger dimension using a neural network A . This is a matrix multiplied by the hidden state at an atom to generate the outgoing message for each atom. The incoming messages for an atom are summed up and fed into the recurrent unit conditioned on the present hidden state of the atom to generate the next hidden state for each atom. The stage is run for T timesteps.

Message passing phase is determined using the message function M_t and vertex update function U_t for vertex b adjacent to x_a :

$$m_v^{t+1} = \sum_{w \in N(v)} M_t(h_v^t, h_w^t, e_{vw})$$

$$h_v^{t+1} = U_t(h_v^t, m_v^{t+1})$$

The message passing itself consists of two parts which are repeated for k steps:

1. The edge network, which passes (a) 1-hop neighbours of w_i v to v , based on the edge features between them in e_{vw} , resulting in updated states of the nodes which are passed to the semi-master node (v') (b) aromatic rings in the molecule to the semi-master node

layer to get aromatic ring feature representation and finally (c) bond feature representation.

2. The gated recurrent network (GRU) takes as input the most recent node state and updates it based on the previous node states of the atoms. The most recent node state serves as input to the GRU, and previous node states are incorporated into the memory state of the GRU. This allows the flow of information from one node state to the other.

At each step $1 \dots k$, the radius or number of hops of aggregated information from v increases by 1 which then passes to the readout stage, and finally, we find the predictions (\hat{y}) for each of the classes.

2.3 Readout Stage

This phase aggregates all the information collected from all nodes and calculates a final value for the epoch.

$$\hat{y} = R(h_a) \text{ for every } x_a \text{ which belongs to the graph } G.$$

R represents the readout stage and h_a represents the hidden state at every time step T. For our network, we use the same message functions, update functions, readout functions, and initial hidden state used by Gilmer et al.¹⁰.

2.4 Semi-Master Nodes

On the existing neural network architecture by Gilmer et al.,¹⁰, we add a new structural layer called Semi-Master Node (Figure 2), which accounts for all the structural units in the molecule

such as different aromatic parts, rings, carboxyl groups among others, which we hypothesize play a significant role in deciding the properties of the molecule during message passing. This unit extracts the atoms features, aromatic ring features, and bond features, which can easily be obtained from the Bond Order Matrix (Figure 1) that shows the spatial information about different aromatic cycles, bond orders and atoms in the molecules (Figure 3).

For example, the dipole moment is a property of a molecule, and structures such as aromatic rings decide the direction of this vector. Similarly, we expected that our change in the neural network would help to either speed up predicting these physicochemical properties with fewer epochs or improve the accuracy with which we could predict these properties.

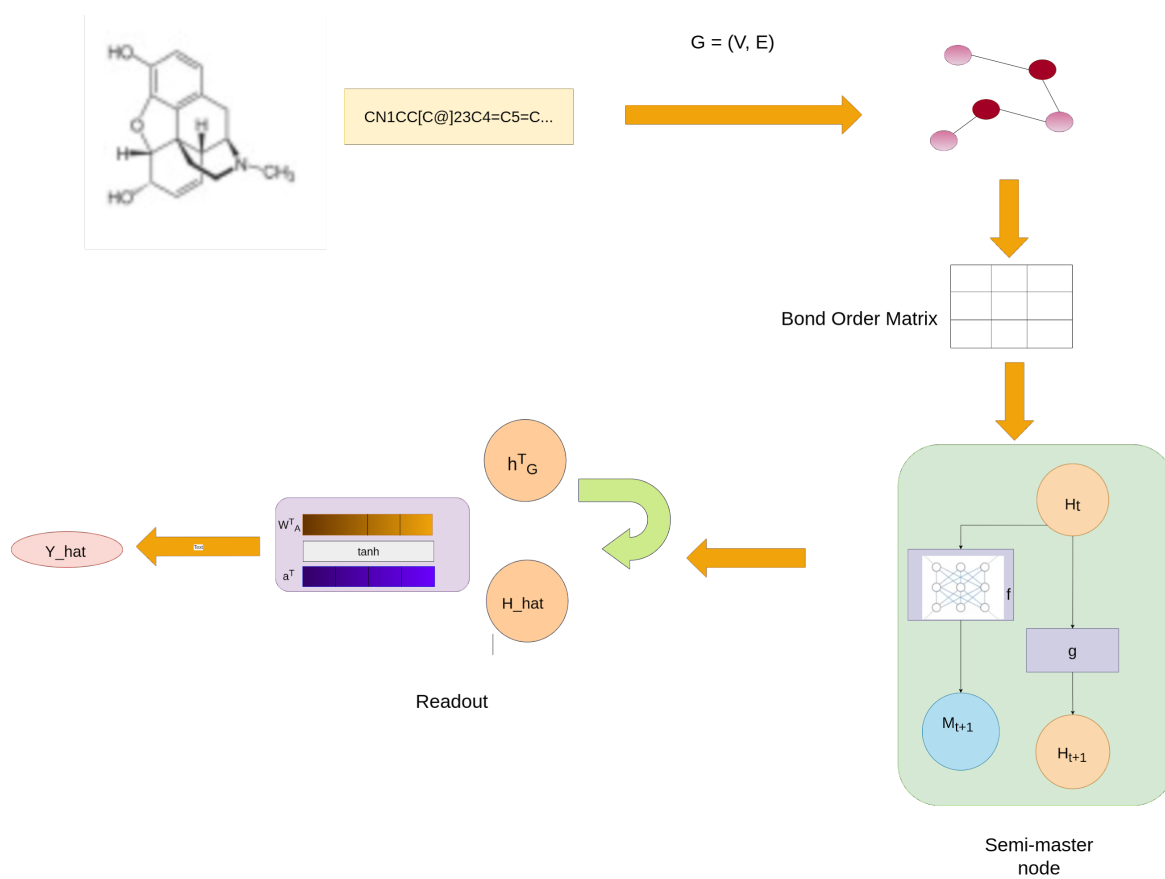


Figure 2. High-level, step-by-step diagram of how our model functions, starting from making a graph, forming a bond order matrix, adding a semi-master node, and running the MPNN, as described in section 2.

3. Model Training

Joint training has been implemented for 40 epochs, with the learning rate being $5e-4$. Along with the MPNN architecture explained above, we are using Adam Optimizer. Batch optimization has been implemented with a batch size of 16. We use Mean Squared Error as the loss function here.

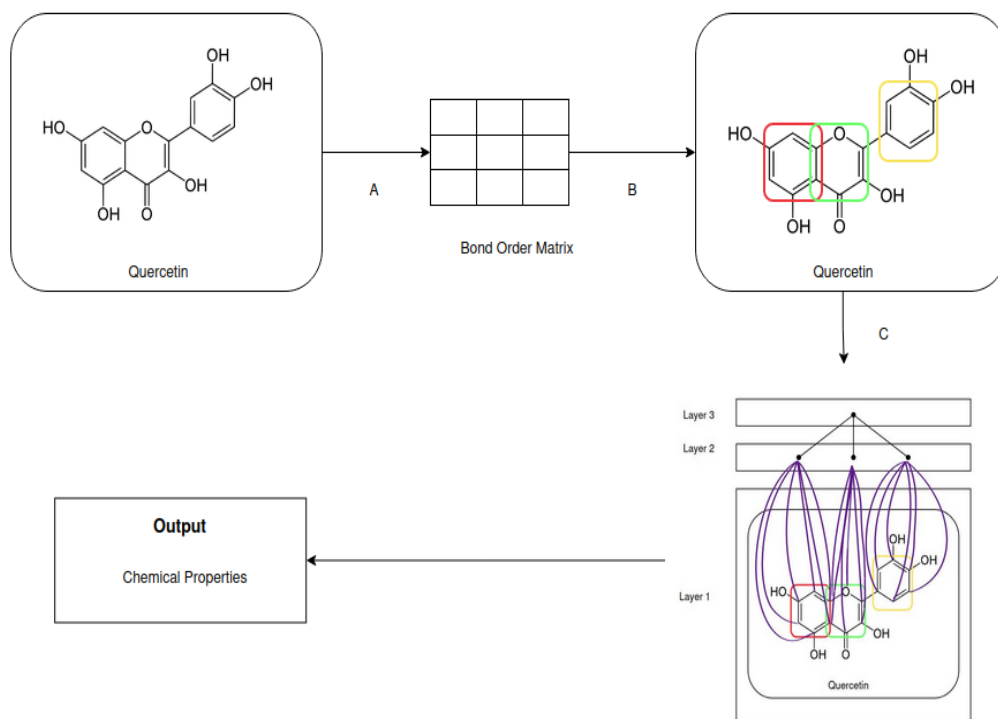


Figure3. An example of the workflow of the presented model. The molecule is converted into its BO matrix and passed through a fully connected layer to get an encoding that captures functional groups. This is then our input to the MPNN framework to generate the properties of the small molecules.

4. QM9 DATASET

To get the physicochemical properties of molecules, Quantum mechanical (QM) calculation is the most accurate way to obtain the properties like total energy, dipole moments, polarizabilities, enthalpies, and free energies of atomization. As a result, the QM9 dataset has become a golden standard for predicting various chemical properties. QM9 consists of 133,885 molecules with up

to nine heavy atoms (only H, C, N, O, and F and up to 9 “heavy” atoms), a subset of the GDB-17 chemical universe of 166.4 billion molecules with up to 17 atoms. All the molecular properties were calculated at the DFT/B3LYP/6-31G(2df,p) level of theory. The table gives a summary of the properties. For the predominant stoichiometry, $C_7H_{10}O_2$, there are 6,095 constitutional isomers among the 134k molecules¹². The properties of each molecule are stored in a file in XYZ format, which is collected and transferred into a CSV file.

No.	Property	Unit	Description
1	<i>tag</i>	—	‘gdb9’ string to facilitate extraction
2	<i>i</i>	—	Consecutive, 1-based integer identifier
3	<i>A</i>	<i>GHz</i>	Rotational constant
4	<i>B</i>	<i>GHz</i>	Rotational constant
5	<i>C</i>	<i>GHz</i>	Rotational constant
6	μ	<i>D</i>	Dipole moment
7	α	α_0^3	Isotropic polarizability
8	$\epsilon HOMO$	<i>Ha</i>	Energy of HOMO
9	$\epsilon LUMO$	<i>Ha</i>	Energy of LUMO
10	<i>egap</i>	<i>Ha</i>	Gap ($\epsilon LUMO - \epsilon HOMO$)
11	$\langle R^2 \rangle$	<i>a02</i>	Electronic spatial extent
12	<i>zpv</i>	<i>Ha</i>	Zero point vibrational energy

13	U_0	Ha	Internal energy at 0 K
14	U	Ha	Internal energy at 298.15 K
15	H	Ha	Enthalpy at 298.15 K
16	G	Ha	Free energy at 298.15 K
17	C_v	$cal/molK$	Heat capacity at 298.15 K

Table1. Description of the physicochemical properties in the QM9 dataset which we are trying to determine.

5. Results and Discussion

Datasets:

In this study, we have focused on the QM9 dataset for predicting properties such as atomization energy, dipole moment, *HOMO – LUMO* gap, enthalpy of formation, and heat capacity at constant pressure, among others, and achieved state-of-the-art results by improving MPNN architecture by adding a semi-master node (Figure 2) layer which acts like an intermediary between the master layer and the subsequent following layers, this helps in handling and coordination of the messages and updates from one node to the other and reduces the burden on the master node hence improving the overall performance and efficiency of the system. We also

include pre-processing methods, including the Bond-Order Matrix. As described in the methods section, MPNNs were initially introduced to reformulate the existing graph neural networks into the Message Passing framework. Formally, a molecular graph is a structure $G = (V, E)$, where V is a set of nodes, each representing one atom, and $E \subset (V \times V)$ is the set of edges, each representing one bond between atoms. Edges may be defined either through pre-existing knowledge about molecular bonds i.e., the Bond Order Matrix (Figure 1) representation of molecules that we introduced above. Properties of an atom $v \in V$, such as the 13 properties of the QM9 dataset in a molecule, are modeled as an f -dimensional node feature vector $x_v \in R_f$ and properties of an edge $v \times w \in E$ are modeled as a d -dimensional edge feature vector $e_{vw} \in R_d$, where attributes of atoms or edges have categorical values. We report the ratio of the mean absolute error (MAE) of our models the provided estimate of chemical accuracy given in Table 5. MAE can be calculated as Error Ratio X Chemical Accuracy Thus, any model with an error ratio of less than 1 has achieved chemical accuracy for that target. The experimental outcomes are presented in tables 2, 3, and 4. In the Appendix, we list the chemical accuracy estimates for each target, these are the same estimates that align with Faber et al.(2017). In this way, the MAE of our models can be computed as the product of Error Ratio and Chemical Accuracy.

Table 2 accounts for the Mean Absolute Error (MAE) ratios for individual properties within the QM9 dataset. Overall, these modifications help on all 13 properties. Still, we observe that 8 properties exhibit AE ratios that are either lower or comparable to those reported in prior methodologies, accompanied by a significantly reduced mean in property errors. This

observation underscores our model's heightened precision and dependability compared to its predecessors.

We have performed several experiments in order to find the best MPNN as well as the proper input representation. It is imperative to highlight that the enhancements made to the MPNN architecture, namely the inclusion of the semi-master node (Figure 2) and the Bond Order matrix (Figure 1), have demonstrably augmented its predictive capacity, as corroborated in Tables 2 and 3. Notably, the synergistic integration of the Bond Order Matrix and Semi-Master Node has resulted in a notable enhancement in our model's performance, leading to a **92.4%** reduction in the Average Error Ratio as compared to the preceding state-of-the-art methodology.

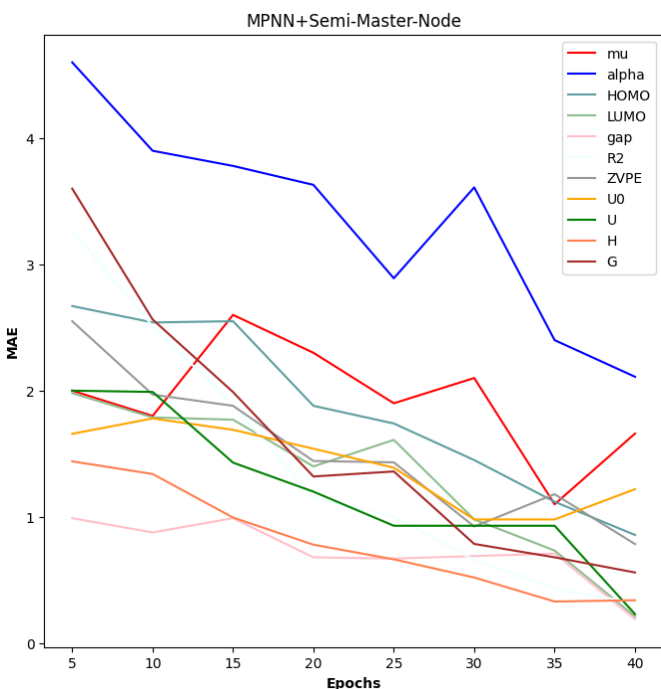


Figure4. The MAE vs Epochs for properties for MPNN with Semi-master Node. We can see a clear reduction in error as the number of epochs run increases. This shows that the semi master node does a good job capturing structural properties over long distances in the molecule.

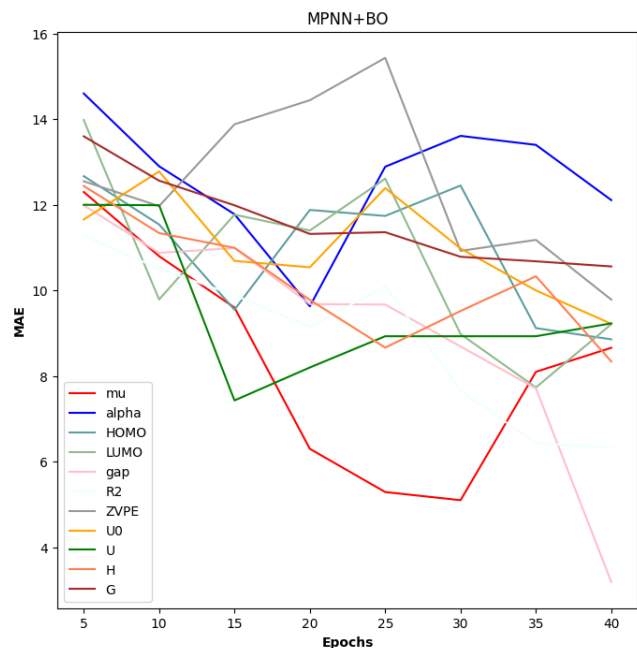


Figure 5. The MAE vs Epochs for properties for MPNN with Bond Order (BO) matrix. There is no clear reduction in the error as epochs run increases. Although BO is a good input representation, by itself it does not achieve much in training the NN.

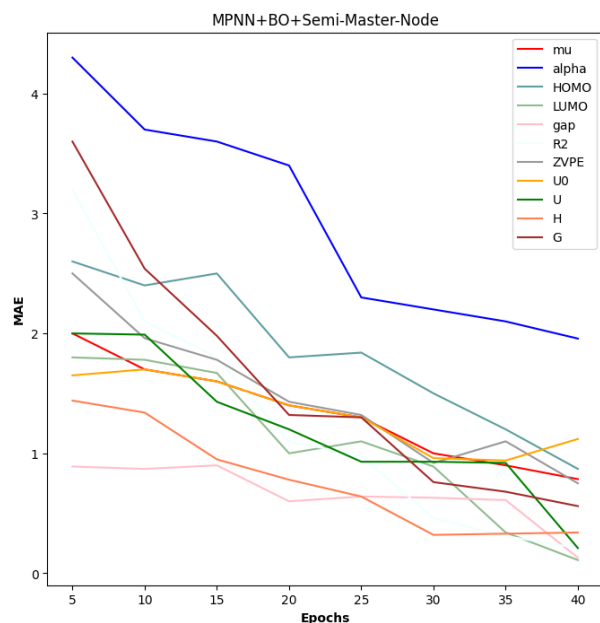


Figure 6. The MAE vs Epochs for properties for MPNN with Bond Order matrix and semi-master node. This graph shows how the combination of BO matrix and semi-master node helps

attain the desired result. The error reduces as the epochs run increases. The advantage here is both in a succinct input representation and the learning rate/accuracy of the semi-master node.

<i>Target</i>	<i>BAML</i>	<i>BOB</i>	<i>CM</i>	<i>ECFP4</i>	<i>HDAD</i>	<i>GC</i>	<i>GG – NN</i>	<i>enn – s2s</i>	<i>enn – s2s – ens5</i>	<i>OurModel</i>
<i>mu</i>	4.34	4.23	4.49	4.82	3.34	0.70	1.22	0.30	0.20	0.7851
<i>alpha</i>	3.01	2.98	4.33	34.54	1.75	2.27	1.55	0.92	0.68	1.957
<i>HOMO</i>	2.20	2.20	3.09	2.89	1.54	1.18	1.17	0.99	0.74	0.87
<i>LUMO</i>	2.76	2.74	4.26	3.10	1.96	1.10	1.08	0.87	0.65	0.11
<i>gap</i>	3.28	3.41	5.32	3.86	2.49	1.78	1.70	1.60	1.23	0.13
<i>R2</i>	3.25	0.80	2.83	90.68	1.35	4.73	3.99	0.15	0.14	0.21
<i>ZPVE</i>	3.31	3.40	4.80	241.54	1.91	9.75	2.52	1.27	1.10	0.75
<i>U0</i>	1.21	1.43	2.98	85.01	0.58	3.02	0.83	0.45	0.33	1.12
<i>U</i>	1.21	1.44	2.99	85.59	0.59	3.16	0.86	0.45	0.34	0.21
<i>H</i>	1.22	1.44	2.99	86.21	0.59	3.19	0.81	0.39	0.30	0.34
<i>G</i>	1.20	1.42	2.97	78.36	0.59	2.95	0.78	0.44	0.34	0.66
<i>Cv</i>	1.64	1.83	2.36	30.29	0.88	1.45	1.19	0.19	0.15	0.19
<i>Avg</i>	2.38	2.27	3.61	62.24	1.46	2.94	1.36	0.67	0.51	0.61

Table2. Property wise MAE ratio for our model and previous approaches

<i>Model*</i>	<i>AverageErrorRatio</i>
<i>GG – NN</i>	3.47
<i>GG – NN + VirtualEdge</i>	2.90

<i>GG – NN + MasterNode</i>	2.62
<i>GG – NN + set2set</i>	2.57
<i>MPNN + SemiMasterNode</i>	1.87

Table3. MAE ratio for MPNN with the addition of semi-master node

<i>Model</i>	<i>Average Error Ratio</i>
<i>GG – NN + jointtraining</i>	1.92
<i>towers8 + jointtraining</i>	1.75
<i>GG – NN + individualtraining</i>	1.53
<i>towers8 + individualtraining</i>	1.37
<i>MPNN + SemiMasterNode + BO + JointTraining(OurModel)</i>	0.61

Table4. MAE ratio for MPNN+semi-master node+Bond Order Matrix

Target	DFT Error	Chemical Accuracy
--------	-----------	-------------------

<i>mu</i>	0.1	0.1
<i>alpha</i>	0.4	0.1
<i>HOMO</i>	2.0	0.043
<i>LUMO</i>	2.6	0.043
<i>gap</i>	1.2	0.043
<i>R2</i>	-	1.2
<i>ZPVE</i>	0.0097	0.0012
<i>U₀</i>	0.1	0.043
<i>U</i>	0.1	0.043
<i>H</i>	0.1	0.043
<i>G</i>	0.1	0.043
<i>C_v</i>	0.34	0.050
<i>Omega</i>	28	10.0

Table 5. Chemical accuracy for each target

Utilization of BO Matrix: The Bond Order matrix is a matrix representation of a molecule where each element of the matrix corresponds to the bond order between two atoms in the molecule. The bond order is typically defined as the number of electron pairs shared between two atoms in a covalent bond.

The bond order matrix provides a more detailed and complete representation of the molecular structure than other common representations such as distance matrix and captures details such as aromatic cycles, atoms, and bonds in the molecule. In particular, the bond order matrix captures

the strength and type of bonds between atoms, which is important for accurately representing molecules' chemical properties and reactivity. Using bond order matrices can also improve the performance of neural networks trained on molecular data. Neural networks are often used in computational chemistry and drug discovery to predict various properties of molecules, such as their activity, solubility, and toxicity. By incorporating bond order matrices as inputs to neural networks, the models can better capture molecules' chemical features and structural characteristics, leading to more accurate predictions. Overall, bond order matrices provide a more comprehensive representation of molecular structure and have the ability to compartmentalize different structures in the molecule as we show in Figure 5. which can be extracted for further feature extractions in the semi-master node, we also see that this can enhance the accuracy and performance of neural networks trained on molecular data as seen in Figure 6 and Table 4, our model which includes using BO matrix and semi-master node has almost 56 % lower errors.

Semi-Master Node:

In MPNNs, messages are passed between nodes in a graph in multiple rounds, with each round allowing nodes to update their state based on information from their neighbours. However, standard MPNNs do not have a mechanism to propagate information over long distances in the graph, which can be important for accurately capturing the global features of the graph.

We introduce the semi-master node as a way to address this limitation by augmenting the MPNN to capture long-range interactions between nodes, which will especially be useful for sparse

graphs. The semi-master node acts as a central node that aggregates information from all other nodes in the graph and feeds it back into the network. This allows the model to capture long-range interactions between nodes and capture global features of the graph.

We show that using a semi-master node has improved the performance of MPNNs molecular property prediction. It has been shown to improve the accuracy of MPNNs on tasks involving molecules, where long-range interactions between atoms can be important for accurately predicting molecular properties.

Owing to this, it made sense to use Neural Networks to generate the output. Considering that today, there is ample data present and enough computational resources to train a model, it does not come as a surprise that individuals would try and shift their approach in that direction. Two prominent approaches by Behler et al⁷ and Rupp et al⁸ have attempted to estimate the solution to quantum mechanics using Neural Networks without directly appealing to DFT. But both approaches used hand-engineered features (e.g.: symmetry functions, Coulomb matrix) that ensured physical aspects were built into their input representations. This is what we avoided in our paper by passing our BO matrix through a semi-master node to generate an encoding that would already constitute the structural information. Additionally, MPNN has been a widely adopted framework for running Neural Network models on graph-based inputs.

6. CONCLUSION

We presented a pre-training strategy for representing molecules in a Bond-Order Matrix format that teaches a GNN to utilize information related to the bonds in the molecules from 2D

molecular graphs. We also present a semi-master node layer that helps capture molecules' important properties in intermediate layers, increasing the model's performance. We found consistently large improvements ($\sim 56\%$) in predicting these properties for different types of molecules. This provides valuable information for downstream applications in molecule design. In our paper, we introduced an innovative pre-training strategy designed to represent molecules in a Bond-Order Matrix format. This approach facilitates the training of MPNN to effectively leverage bond-related information derived from 2D molecular graphs. Additionally, we proposed the incorporation of a semi-master node layer within the network architecture. This semi-master node layer was specifically designed to capture crucial molecular properties during intermediate layers, thereby enhancing the overall performance of the model. Our experiments yielded consistently significant improvements, with a notable enhancement observed in quantum properties across various types of molecules. This novel methodology demonstrates the potential of combining pre-training strategies and specialized network architectures to achieve substantial advancements in predicting key molecular properties.

7. DATA AND SOFTWARE AVAILABILITY STATEMENT

The dataset is available at <http://quantum-machine.org/datasets/>.

8. ACKNOWLEDGEMENTS

RA acknowledges the funding from DBT-NNP (Grant No. BT/PR40236/BTIS/137/51/2022).

The authors thank Prof. Ashwin Srinivasan and Prof. Snehanshu Saha for their valuable discussions. RA is an affiliated faculty member of APPCAIR, BITS Pilani, Goa Campus. The

authors also would like to acknowledge Kosambi, High Performance Computing at BITS Pilani, Goa Campus for computing resources.

9. REFERENCES

1. Lipinski, C. A., Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies* **2004**, *1* (4), 337-341.
2. (a) Roskoski, R., Rule of five violations among the FDA-approved small molecule protein kinase inhibitors. *Pharmacological Research* **2023**, *191*, 106774; (b) Zhang, M.-Q.; Wilkinson, B., Drug discovery beyond the 'rule-of-five'. *Current Opinion in Biotechnology* **2007**, *18* (6), 478-488.
3. Doak, B. C.; Kihlberg, J., Drug discovery beyond the rule of 5 - Opportunities and challenges. *Expert Opinion on Drug Discovery* **2017**, *12* (2), 115-119.
4. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98* (7), 5648-5652.
5. Hedin, L., New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem. *Physical Review* **1965**, *139* (3A), A796-A823.
6. Ceperley, D.; Alder, B., Quantum Monte Carlo. *Science* **1986**, *231* (4738), 555-560.
7. Behler, J.; Parrinello, M., Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters* **2007**, *98* (14), 146401.
8. Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A., Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Physical Review Letters* **2012**, *108* (5), 058301.
9. Niepert, M.; Ahmed, M.; Kutzkov, K., Learning convolutional neural networks for graphs. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, JMLR.org: New York, NY, USA, 2016; pp 2014–2023.
10. 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*, Doina, P.; Yee Whye, T., Eds. PMLR: Proceedings of Machine Learning Research, 2017; Vol. 70, pp 1263--1272.
11. Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; Monfardini, G., The Graph Neural Network Model. *IEEE Transactions on Neural Networks* **2009**, *20* (1), 61-80.
12. (a) Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L., Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *Journal of Chemical Information and Modeling* **2012**, *52* (11), 2864-2875; (b) Ramakrishnan, R.; Dral, P.

O.; Rupp, M.; von Lilienfeld, O. A., Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data* **2014**, *1* (1), 140022.