Stereoelectronics-Aware Molecular Representation Learning

Daniil Boiko[†]
Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
dboiko@andrew.cmu.edu

Thiago Reschützegger[†]
Department of Chemical Engineering
Federal University of Santa Maria
Santa Maria, RS
thiago.reschutzegger@acad.ufsm.br

Benjamin Sanchez-Lengeling
Google Research
bmsanchez@google.com

Samuel M. Blau
Energy Technologies Area
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
smblau@lbl.gov

Gabriel dos Passos Gomes[*]
Department of Chemistry
Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
gabegomes@cmu.edu

Abstract

The representation of molecular structures is crucial for molecular machine learning strategies. Although graph representations are highly versatile and show their broad applicability, they lack information about the quantum-chemical properties of molecular structures. This work proposes a new way to infuse such information into molecular graphs, using a supervised learning method. As a result, the model is able to predict essential higher-order interactions between electron-rich and electron-deficient localized orbitals. The learned interactions are then used as a representation for the prediction of downstream tasks, improving over QM9 baselines.

1 Introduction

The prediction of molecular properties is at the core of chemical, biological, and material sciences. From the discovery of materials for solar panels[1] to the record-setting development of a new drug,[2] molecular machine learning (ML) greatly impacted modern science by enabling fast inference. The performance of ML methods in these fields is strongly connected to the molecular representation design, being (arguably) the most important factor for its success. Moreover, with the increasingly larger amount of available chemical data, many methods were developed to learn the best representation in a data-driven fashion. Especially, graph neural networks (GNNs) have been the protagonists of these advances by leveraging the molecular topology to correlate structure with the desired global property.[3]

Despite the success of machine learning for molecular property predictions, the representations used in these methods are incomplete. In particular, current graph representations lack quantum-chemical

Preprint. Under review.

priors from the electronic structure of molecules. Such considerations are essential for understanding the fundamentals of molecular interactions, chemical reactivity, and property predictions. Even though these electronic features can be learned implicitly by molecular geometry, such a process requires large and expensive datasets.[4, 5] Moreover, there would be no guarantee that the pre-trained model would preserve all required information for downstream tasks.

The computational chemistry community has developed solid foundations for describing the quantum-mechanical nature of molecular structures. For example, it is well-known that molecules are not only composed of covalent interactions but also stabilizing noncovalent interactions between the corresponding bond and lone pair orbitals. However, these non-trivial interactions are not currently considered when designing computational representations of molecules, even though they play a crucial role in describing a wide range of properties. These stabilizing donor-acceptor interactions are essential in our understanding of many chemical phenomena and have been disregarded by the machine learning community.

This work proposes a new approach to representing molecular structures with graphs that capture orbital interactions. We extend standard molecular structure graphs with additional nodes representing bonding and non-boning orbitals. Such information is then used to predict intra-molecular interactions that are not explicitly given by the original topology. This work describes learning features for these nodes from quantum chemistry calculations and extending original graphs improves over baselines.

2 Related work

Molecular representation learning In recent years, the field has greatly benefited from many novel representations of molecules in data-driven approaches. Ways to represent molecules include hand-crafted molecular descriptors/features,[6–11] language representations,[12–15] and graphs. Graph-based representations achieved state-of-the-art results by employing inherent topological structures of molecules in neural networks, known as Graph Neural Networks (GNNs).[5, 16–22] Moreover, recent advances combine GNNs with tridimensional geometry into roto-translation equivariant representations, increasing the amount of information to perform correlations.[23–27]

Graph augmentation Previous works have reported the difficulty of GNNs to propagate information across distant nodes. Such phenomena are often referred to as *over-squashing*, a major problem when node properties depend on long-range interactions.[28, 29] In molecular settings, these long-range interactions are often caused by noncovalent orbital interactions that are not explicitly given by the standard molecular topology. Several methods to mitigate this problem have been proposed, such as graph augmentations,[30] fully-connected graph convolutional layers,[28] and connectivity-independent long-range reasoning.[31] To address this problem, we present a graph augmentation method grounded on electronic interactions.

Electronic information on graphs Across the literature, physical and chemical priors have been successfully included in data-driven models. Usually, these methods use hand-crafted features from chemistry domain knowledge, showing major improvements over baselines.[32–34] However, a recent effort has been made to model quantum interactions automatically within a molecule.[35–37] In this work, we propose a molecular electronically-aware molecular featurization method that can be easily combined with existing representations.

Natural localized orbitals At a high level, the natural bond orbital (NBO) analysis provides electron density information, providing a valid description of a molecule's wave function. Instead of representing these densities as delocalized molecular orbitals, this method yields a collection of localized natural atomic orbitals, hybrid orbitals, and bond orbitals.[38] Moreover, NBO analysis expresses interactions between filled orbitals (donors) and empty orbitals (acceptors).

3 Graph representation of molecular electronic properties

Modeling non-bonding orbitals is a non-trivial task since such information is not accessible from traditional molecular graphs. To solve it, we formulated the problem as a sequential graph construction model. We accomplished the ultimate goal by training this model using NBO analysis data via

supervised learning. As a result, the full pipeline can predict NBO interactions based on complex electronic interactions within the molecule.

Firstly, the molecular graph $\mathcal{G}_M$ is constructed from the three-dimensional input structure. After that, the graph is used as an input for the first model, which is responsible for predicting the number of lone pairs and their types. Once the model has successfully predicted lone pair information, we store it in the extended molecular graph $\mathcal{G}_{ext}$. Also, we propose the addition of nodes that trivially represent σ- and π-bonds between two atoms. We argue that this is a natural way to describe connections in molecular settings instead of simply representing bonds as edges.

The extended molecular graphs are sequentially used as input for a second model, designed to predict electronic properties of orbital overlaps in a multitask fashion. Therefore, the model can predict the interaction between combinations of lone pairs and bonds, called second-order interactions.

We describe the graph of electronic properties as a graph $\mathcal{G}_E = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ are nodes of the graph, and $\mathcal{E}$ are the edges. In this case, $\mathcal{V}$ can represent atoms, lone pairs, and bond orbitals, where a mapping function $\phi = \mathcal{V} \rightarrow \mathcal{A}$ describes the types of nodes in the graph. Edges of this graph can represent atom to atom covalent interactions, node to bond, and second-order interactions between possible combinations of lone pairs and bond orbitals.

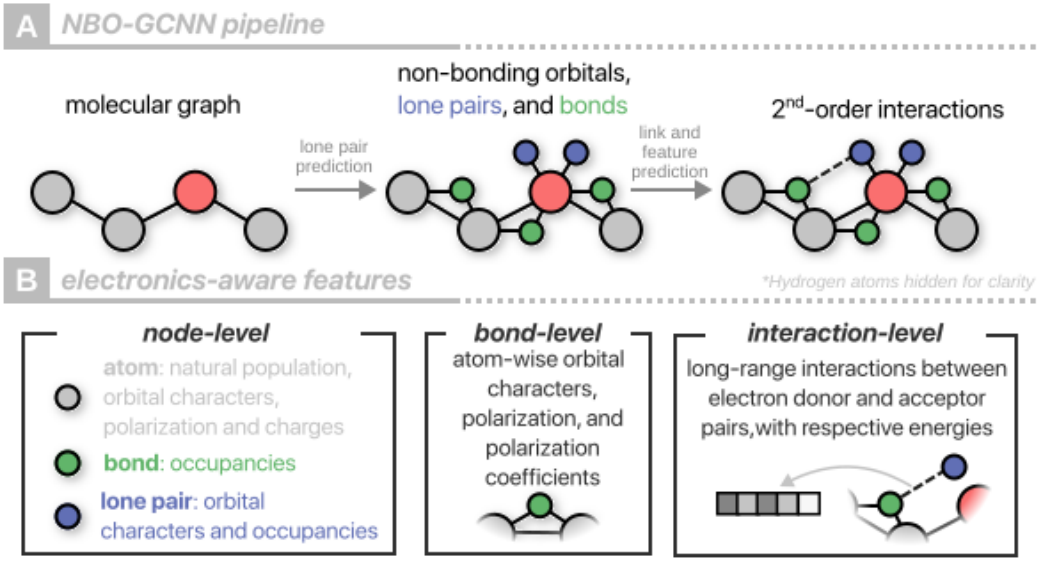

Figure 1: **Overview of our approach. a. Scheme of the pipeline** We train models to sequentially extend the graph with non-bonding orbitals followed by its interactions. **b. Construction of electronics-aware features for downstream tasks** The final model is trained to predict *node*-, *bond*-, and *interaction*-level electronic features, together with donor-acceptor interaction predictions.

3.1 Data collection

Given the optimized structures for each molecule of the QM9 dataset,[39] we conducted single-point calculations at the def2-TZVPD[40, 41] + ωB97M-V[42] level of theory. We used Q-Chem 5.4.2[43] interfaced with NBO5[44] to perform the calculations resulting in targets for atom, bond, lone pairs, and orbital interactions, illustrated in Figure 1b. These are described as follows:

Atom features The performed natural atomic orbital analysis returns localized electron information for each atom. Atom targets include its charge, the number of core electrons, valence electrons, and total electrons. Even though NBO analysis provides Rydberg orbitals, we did not keep them as a model's target due to the controversy in the physical meaning.

Bond features In the context of localized natural bond orbitals, bonds are simply a combination of the orbitals from each atom. For that reason, the NBO analysis data provides atom-wise *s*, *p*, *d*, and *f* characters, polarization, polarization coefficient, and the respective values for antibonding orbitals. Occupancy for bonding and antibonding orbitals are the only bond-specific target from the original data. In total, they are totalizing twenty-six targets.

Lone pair features Orbital hybridization is described by the *s*, *p*, *d*, and *f* characters. Also, the NBO analysis provides information about its occupancy, summing up to five targets.

Orbital (2nd order) interactions These represent the interactions between donor and acceptor orbitals. Donors, represented by lone pairs n, σ, and π bonds, are electron-rich orbitals, while acceptors, represented by σ^*, and π^* anti-bonds are electron deficient. In practical terms, our ground truth graph represents one donor-acceptor interaction as a connection between the respective nodes. The NBO analysis quantified these interactions by the perturbation energy, energy difference, and the Fock matrix element, totalizing three targets.

3.2 Input graph construction

Following the pipeline showcased in Figure 1, the first step is to construct homogeneous molecular graphs from the input structure. Molecular graphs $\mathcal{G}_M = (\mathcal{V}, \mathcal{E})$ consist of a set of nodes $\mathcal{V}$ representing atoms, and edges $\mathcal{E}$, representing covalent bonds. The community has widely explored this representation, achieving excellent results in property prediction tasks when combined with GNNs. In this work, the molecular graph $\mathcal{G}_M$ is comprised by a set of atoms $u \in \mathcal{V}$, encoded by the respective feature vector x_u (i.e., the element), along with bonds $(u, v) \in \mathcal{E}$ encoded by a feature vector $x_{u,v}$ (i.e., the bond order and distance).

3.3 Lone pair prediction

Tasks Even though many heuristics can be used to define the number of lone pairs, we argue that a data-driven model is more well-suited in this case since it can interpolate between different contexts. With this in mind, we built a neural network capable of predicting the number of lone pairs for each atom and their types. Such types were used to distinguish lone pairs of the same atom since there are possible differences in their NBO data. For that reason, we determined the types by an analytical threshold relating *s*- and *p*-characters, expressing the conjugation likelihood of a given lone pair. The threshold is defined by the equation below:

$$p\text{-character} - s\text{-character} > 80$$

Graph encoder Both tasks were tackled in tandem with a mapping function $f : \mathcal{G}_M \rightarrow \mathcal{G}_{ext}$, modeled as a GNN. To mitigate the over-smoothing problem, we used multiple aggregation functions through the message-passing scheme,[45] along with residual connections.[46, 47] The encoder is constructed by stacking several propagation layers followed by a ReLU activation function. Node embeddings are then concatenated with a residual connection from the input graph. The result is forwarded to a multilayer perceptron (MLP), which in this case, is composed of two linear layers separated by a ReLU activation layer.

Training As a design choice, we framed both tasks as node-level classifications, in which each class represents the number of lone pairs, and how many satisfy the threshold, respectively. Therefore, we used the sum of the cross-entropy loss of each task as the loss function.

Extended molecular graph construction Given the predicted number of lone pairs and their types per atom, the pipeline continues toward the construction of the extended molecular graph $\mathcal{G}_{ext}$. As previously mentioned, the nodes in the extended graph do not represent only atoms but also lone pairs and bonds. Even though these two nodes do not represent atoms, we opted to design the graph as homogeneous rather than heterogeneous. Next, predicted lone pair nodes were connected to the respective atom node, carrying features such as its type (σ- or π-), and distance from the parent atom. Finally, we created the bond nodes representing different bonds between two atoms. Each pair of atoms is connected to the number of bond nodes corresponding to the bond order.

3.4 Learning natural bond interactions

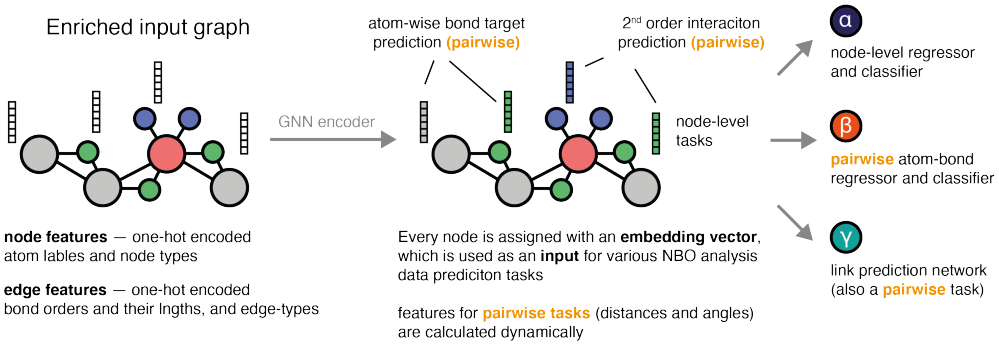

Figure 2: Description of the approach for the prediction of NBO analysis data

The neural network architecture comprises two parts (shown in Figure 3): the node encoder and a group of multiple separate MLPs. The latter make predictions using the embeddings from the encoder part, but multiple preprocessing steps may be conducted depending on the task.

The encoder is constructed by stacking multiple graph neural network blocks and concatenating the outputs of each block. A block comprises a graph attention layer and a ReLU activation layer. None of the dropout or batch normalization layers was used.[19] Concatenated outputs were then passed into the MLP network with one single layer to construct the node embedding.

This encoder architecture is designed to tackle the over-smoothing issues of graph neural networks.[48] The problem arises when multiple graph neural network layers are stacked together, making the computational graphs nearly identical. Over-smoothing might not be a severe issue for graph-level tasks but is a significant problem in performing node-level tasks. Multiple solutions were proposed: residual connections augmentations.[46, 47] In this work, we concatenate outputs of intermediate layers tackling both the over smoothing and vanishing gradients problems.

All MLPs for separate tasks follow the same architecture: one linear layer, the ReLU activation, the batch normalization layer, and one final linear layer. The following sections describe input preparation and loss function for these networks.

$$\mathcal{L} = \mathcal{L}_\alpha + \mathcal{L}_\beta + \mathcal{L}_\gamma + \mathcal{L}_\delta$$

Atom, lone pair, and bond nodes The most straightforward problem to solve is the prediction of targets for individual nodes. Here, for all types of nodes, only one network is used. The loss function was defined as a sum of separate losses for each node type. For all features except orbital characters of lone pairs, the mean squared error (MSE) loss was used. Orbital character prediction was optimized with a cross-entropy loss function. MSE was also the key metric for this type of task. R^2 scores were also recorded.

$$\mathcal{L}_\alpha = \text{MSE}(\alpha(x), y) + \text{BCE}(\alpha(x), y)$$

Atom-wise bond target prediction Some of the bond features are related to each of the atoms. To keep permutation invariance, it was impossible to predict them in the previous step. The task was solved by concatenating embeddings of the atom in question and the corresponding bond and then passing it into the MLP. Polarization values prediction was optimized with MSE loss. Orbital characters were optimized with cross-entropy loss. Similar to the previous section, MSE and the R^2 score were used to control the training.

$$\mathcal{L}_\beta = \text{MSE}(\beta(x), y) + \text{BCE}(\beta(x), y)$$

Link prediction approach Orbital interactions data is not available directly from the molecular structure, so it should be predicted first. Therefore, the problem was formulated as a link prediction task, which is essentially a classification problem. “Positive examples” (i.e., cases where there is an

Table 1: Results in downstream tasks

	precision	recall	F1 score	% C hits	% H hits	% N hits	% O hits	% F hits
Qtd. of LPs	1.00	0.98	0.99	99.92	100	99.43	99.18	100
Qtd. of LPs within threshold	1.00	0.98	0.99	99.99	97.05	99.92	99.15	100

interaction) were taken from the original dataset, while "negative examples" were sampled from other possible combinations of bonds and lone pairs. Moreover, the direction is essential as (in our case) it describes the donor-acceptor pair but not vice versa. Input data consisted of concatenated embeddings of corresponding nodes and dynamically calculated pairwise features. We performed the training with binary cross-entropy loss. Standard classification metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC AUC) were calculated.

$$\mathcal{L}_\gamma = \text{BCE}(\gamma(x), y) = \sum_{i,j}^{n} -y \log(\gamma(x_i, x_j, p_{i,j})) - (1-y) \log(1 - \gamma(x_i, x_j, p_{i,j}))$$

Interaction edge target prediction Predicted interactions can then be used to input the network for interaction target prediction. The features were also obtained by concatenating node embeddings and dynamically calculated pairwise features. Finally, we trained the network with MSE loss.

$$\mathcal{L}_\delta = \text{MSE}(\delta(x), y)$$

4 Experiments

4.1 Lone pair prediction

The model described in Section 3.3 was trained based on the lone pair properties predicted with NBO5.[44] Based on statistical metrics on the test set shown in Table 1, we observe great results for both tasks. First, we report the macro average precision, recall, and F1 score, followed by the percentage of hits per element. Moreover, the model is able to construct the ground-truth extended graph 98.13% of the cases.

4.2 NBO analysis data prediction performance

The described model was tested in the prediction of NBO analysis data calculated for the QM9 dataset. As it can be seen from Table 2, good or excellent results are obtained for most tasks.

Node-level tasks included predicting properties for atoms, lone pairs, and bonds. For atoms, excellent R^2 scores were obtained; MAEs and RMSEs were less than 0.03 electrons. Lone pair-related tasks included *s*-, *p*-, *d*-, *f*-characters, and occupancies. For *s*- and *p*-character prediction tasks excellent scores were obtained. Results for *d*- and *f*-character prediction were worse, but still very good. Occupancy prediction was also successful. For bonds, occupancies were predicted with good scores. Differences in hybridization characters and polarizations were also successful except *d*- and *f*-characters, which have minimal applicability here, as only the first two row elements were considered. It is important to note that the prediction of difference in *s*- and *d*- characters works better than the prediction of separate values (R^2 of 0.999 vs. 0.997, see below).

Some bond targets are related to consisting atoms; this includes hybridization characters and polarisations, and polarization coefficients. These were predicted with a separate network, which used corresponding node embedding as an input. As a result, excellent performance was achieved for all targets except *d* and *f* characters for the previously discussed reasons.

Prediction of second-order interactions is a classic link prediction problem. Standard regression metrics do not properly represent the model's performance, so we used ROC AUC (area under the receiver operating characteristic curve) and mean ROC AUC. The value of ROC AUC is equal to 0.9934, which is a good result for such complex tasks. Interestingly, the value for the mean ROC AUC of 0.9926 is a bit smaller, suggesting that there might be some outlier molecules. Finally, the

properties of these interactions were also predicted. As can be seen from the data, the only drop in performance is observed for the values of Fock-matrix elements.

As a result, the described GNN architecture is fully capable of predicting NBO data. Notably, the model shows much higher performance than standard NBO analysis calculations and does not have any limit on the number of basis orbitals. Therefore, it opens new opportunities in the analysis of very large molecules and materials.

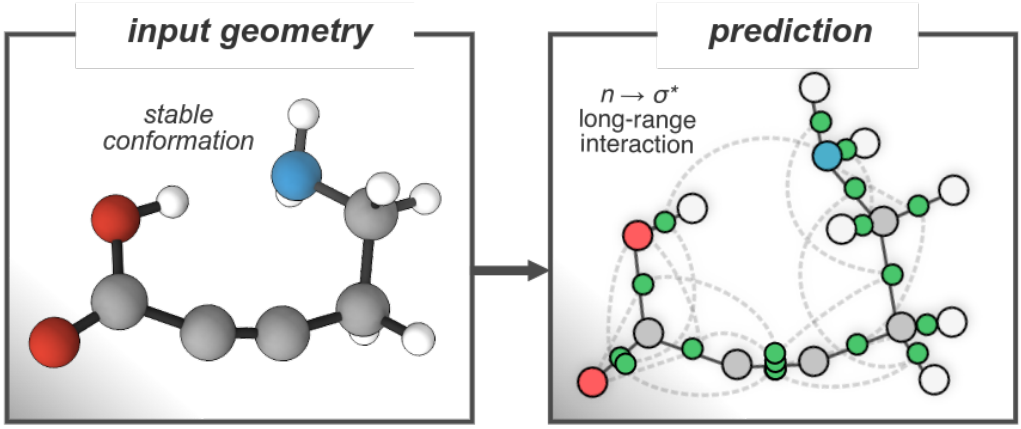

Figure 3: **Cherry-picked example.** The model is able to capture a long-range interaction between an electron-rich orbital (lone pair) and an electron-deficient orbital (sigma anti-bond).

4.3 Performance in downstream tasks

The last step of the work is the analysis of model performance in downstream tasks. To measure the effect of graph augmentation, the following graph structures were compared: standard molecular graphs, molecular graphs with added lone pair and bond nodes (including edges of interactions between them), and molecular graphs with added lone pair and bond nodes with corresponding features, derived from NBO analysis data. Comparing this way, we separated effects from additional features and increased information flow in the graph.

Results in Table 3 clearly show the superior performance of the described approach in the molecular property prediction task. However, changes in graph topology alone also improve the model's performance.

5 Conclusions and future work

Limitations The described approach highly depends on the training data. In particular, elements are one-hot encoded, so to add another element, one needs to collect an extended dataset of NBO analysis data, including the new element. This could be circumvented by using the physical properties of elements as features, but this requires further research. Also, the scope of this work was limited to predicting features of *s*- and *p*- orbitals, therefore not being able to model *d*- and *f*- orbitals accurately.

Broader impact Molecular machine learning is a critical component of pipelines for drug and material discovery, catalyst optimization, and a valuable tool for studying complex biochemical processes. Infusion of electronic data into graph representations for molecular ML will increase trust in these algorithms, contribute to increased interpretability of the models, and open new opportunities to research the relationship between electronic structure and molecular properties. Moreover, this work can also be used for the theoretical chemistry community once it allows high-throughput NBO analysis. The predicted orbitals can be applied to analyze chemical reactivity in a wide range of systems.

Conclusions A novel approach for graph augmentation with NBO analysis-derived data was developed. In the first step, molecular graphs are extended with bond nodes and lone pair nodes, the number of which is computed by a graph neural network. The extended graphs are then used as an

Table 2: NBO analysis data prediction performance

prediction task	RMSE	MAE	R^2	ROC AUC
atom				
charge	0.009	0.006	0.999	
core	0.009	0.008	1.000	
valence	0.021	0.017	1.000	
total	0.028	0.025	1.000	
lone pair				
s	0.008	0.005	0.999	
p	0.008	0.005	0.999	
d	0.000	0.000	0.961	
f	0.000	0.000	0.948	
occupancy	0.008	0.003	0.991	
bond				
occupancy	0.008	0.007	0.907	
s diff.	0.008	0.006	0.999	
p diff.	0.009	0.006	0.999	
d diff.	0.002	0.001	0.307	
f diff.	0.002	0.002	-117.975	
polarization diff.	0.010	0.006	0.993	
polarization coefficient diff.	0.008	0.005	0.991	
atom-wise bond targets				
s	0.017	0.005	0.997	
p	0.017	0.005	0.997	
d	0.002	0.002	-0.494	
f	0.002	0.002	-129.689	
polarization	0.006	0.004	0.997	
polarization coefficient	0.006	0.004	0.994	
link prediction	0.126	0.022	0.861	0.9934
interaction targets				
perturbation energy	0.019	0.006	0.943	
energy difference	0.029	0.010	0.985	
fock matrix element	0.008	0.005	0.899	

Table 3: Results in downstream tasks

	mu	alpha	homo	lumo	gap	r2	zvpe	cv
Standrd graphs	4.00 ± 0.07	8.04 ± 0.37	2.05 ± 0.01	2.30 ± 0.05	2.93 ± 0.68	29.49 ± 2.13	42.61 ± 2.13	7.22 ± 0.21
+ LP and bond	3.77	4.78	1.83	1.95	2.57	22.21	25.45	4.55
+ NBO features	**3.27**	**3.94**	**1.56**	**1.67**	**2.22**	**17.43**	**22.07**	**3.18**

input for another model, which solved a multitask NBO analysis data prediction problem (node-level and link prediction tasks). Finally, the extended graphs are enriched with features that can be used to represent various downstream tasks. As an example, we show increased performance over baselines in predicting the properties of molecules from the QM9 dataset.

Acknowledgements

The authors thank XSEDE, Google Cloud Platform, and the Ulissi group for the provided computing resources.

References

[1] Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, David Duvenaud, Dougal Maclaurin, Martin A. Blood-Forsythe, et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Mater*, 15(10):1120–1127, October 2016. Number: 10 Publisher: Nature Publishing Group.

[2] Alex Zhavoronkov, Yan A. Ivanenkov, Alex Aliper, Mark S. Veselov, Vladimir A. Aladinskiy, Anastasiya V. Aladinskaya, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol*, 37(9):1038–1040, September 2019. Number: 9 Publisher: Nature Publishing Group.

[3] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, January 2020.

[4] J. S. Smith, O. Isayev, and A. E. Roitberg. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.*, 8:3192–3203, 2017.

[5] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1263–1272. PMLR, July 2017.

[6] Hirotomo Moriwaki, Yu-Shi Tian, Norihito Kawashita, and Tatsuya Takagi. Mordred: a molecular descriptor calculator. *Journal of Cheminformatics*, 10(1):4, February 2018.

[7] Sabrina Jaeger, Simone Fulle, and Samo Turk. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J. Chem. Inf. Model.*, 58(1):27–35, January 2018.

[8] David Rogers and Mathew Hahn. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.*, 50(5):742–754, May 2010.

[9] Joseph L. Durant, Burton A. Leland, Douglas R. Henry, and James G. Nourse. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci*, 42(6):1273–1280, December 2002.

[10] Alpha A. Lee, Qingyi Yang, Asser Bassyouni, Christopher R. Butler, Xinjun Hou, Stephen Jenkinson, and David A. Price. Ligand biological activity predicted by cleaning positive and negative chemical correlations. *PNAS*, 116(9):3373–3378, February 2019.

[11] Andreas Mayr, Günter Klambauer, Thomas Unterthiner, Marvin Steijaert, Jörg K. Wegner, Hugo Ceulemans, Djork-Arné Clevert, and Sepp Hochreiter. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chem. Sci.*, 9(24):5441–5451, June 2018.

[12] David Weininger. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, 28(1):31–36, February 1988.

[13] Philippe Schwaller, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A. Hunter, Costas Bekas, and Alpha A. Lee. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.*, 5(9):1572–1583, September 2019.

[14] Garrett B. Goh, Nathan O. Hodas, Charles Siegel, and Abhinav Vishnu. SMILES2Vec: An Interpretable General-Purpose Deep Neural Network for Predicting Chemical Properties. *arXiv:1712.02034 [cs, stat]*, March 2018. arXiv: 1712.02034.

[15] Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alán Aspuru-Guzik. Self-Referencing Embedded Strings (SELFIES): A 100% robust molecular string representation. *Mach. Learn.: Sci. Technol.*, 1(4):045024, December 2020. arXiv: 1905.13741.

[16] Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, et al. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.*, 59(8):3370–3388, August 2019.

[17] Connor W. Coley, Regina Barzilay, William H. Green, Tommi S. Jaakkola, and Klavs F. Jensen. Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction. *J. Chem. Inf. Model.*, 57(8):1757–1772, August 2017.

[18] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv:1609.02907 [cs, stat]*, February 2017. arXiv: 1609.02907.

[19] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *arXiv:1710.10903 [cs, stat]*, February 2018. arXiv: 1710.10903.

[20] Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, et al. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *J. Med. Chem.*, 63(16):8749–8760, August 2020.

[21] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des*, 30(8):595–608, August 2016.

[22] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alan Aspuru-Guzik, and Ryan P Adams. Convolutional Networks on Graphs for Learning Molecular Fingerprints. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[23] Xiaomin Fang, Lihang Liu, Jieqiong Lei, Donglong He, Shanzhuo Zhang, Jingbo Zhou, Fan Wang, Hua Wu, and Haifeng Wang. Geometry-enhanced molecular representation learning for property prediction. *Nat Mach Intell*, 4(2):127–134, February 2022.

[24] Kenneth Atz, Francesca Grisoni, and Gisbert Schneider. Geometric deep learning on molecular representations. *Nat Mach Intell*, 3(12):1023–1032, December 2021.

[25] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv:1802.08219 [cs]*, May 2018. arXiv: 1802.08219.

[26] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. *Advances in Neural Information Processing Systems*, 33:1970–1981, 2020.

[27] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) Equivariant Graph Neural Networks. *arXiv:2102.09844 [cs, stat]*, February 2022. arXiv: 2102.09844.

[28] Uri Alon and Eran Yahav. On the Bottleneck of Graph Neural Networks and its Practical Implications. *arXiv:2006.05205 [cs, stat]*, March 2021. arXiv: 2006.05205.

[29] Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. *arXiv:2111.14522 [cs, stat]*, March 2022. arXiv: 2111.14522.

[30] Rickard Brüel-Gabrielsson, Mikhail Yurochkin, and Justin Solomon. Rewiring with Positional Encodings for Graph Neural Networks. *arXiv:2201.12674 [cs]*, February 2022. arXiv: 2201.12674.

[31] Zhanghao Wu, Paras Jain, Matthew A. Wright, Azalia Mirhoseini, Joseph E. Gonzalez, and Ion Stoica. Representing Long-Range Context for Graph Neural Networks with Global Attention. *arXiv:2201.08821 [cs]*, January 2022. arXiv: 2201.08821.

[32] K. Hansen, Franziska Biegler, R. Ramakrishnan, Wiktor Pronobis, O. A. von Lilienfeld, K. Müller, and A. Tkatchenko. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *The journal of physical chemistry letters*, 2015.

[33] Grégoire Montavon, Katja Hansen, Siamac Fazli, Matthias Rupp, Franziska Biegler, Andreas Ziehe, et al. Learning Invariant Representations of Molecules for Atomization Energy Prediction. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[34] B. Huang and O. A. von Lilienfeld. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *The Journal of chemical physics*, 2016.

[35] Chengqiang Lu, Qi Liu, Chao Wang, Zhenya Huang, Peize Lin, and Lixin He. Molecular Property Prediction: A Multilevel Quantum Interactions Modeling Perspective. *arXiv:1906.11081 [physics]*, June 2019. arXiv: 1906.11081.

[36] Zeren Shui and G. Karypis. Heterogeneous Molecular Graph Neural Networks for Predicting Molecule Properties. *2020 IEEE International Conference on Data Mining (ICDM)*, 2020.

[37] Jeonghee Jo, Bumju Kwak, Byunghan Lee, and Sungroh Yoon. Flexible dual-branched message passing neural network for quantum mechanical property prediction with molecular conformation. *ArXiv*, 2021.

[38] Igor V. Alabugin. *Stereoelectronic Effects: A Bridge Between Structure and Reactivity*. John Wiley & Sons, Ltd, Chichester, UK, September 2016.

[39] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Sci Data*, 1(1):140022, December 2014.

[40] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7(18):3297–3305, August 2005.

[41] Dmitrij Rappoport and Filipp Furche. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.*, 133(13):134105, October 2010.

[42] Narbe Mardirossian and Martin Head-Gordon. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.*, 144(21):214110, June 2016. Publisher: American Institute of Physics.

[43] Evgeny Epifanovsky, Andrew T. B. Gilbert, Xintian Feng, Joonho Lee, Yuezhi Mao, Narbe Mardirossian, Pavel Pokhilko, Alec F. White, Marc P. Coons, Adrian L. Dempwolff, Zhengting Gan, Diptarka Hait, Paul R. Horn, Leif D. Jacobson, Ilya Kaliman, Jörg Kussmann, Adrian W. Lange, Ka Un Lao, Daniel S. Levine, Jie Liu, Simon C. McKenzie, Adrian F. Morrison, Kaushik D. Nanda, Felix Plasser, Dirk R. Rehn, Marta L. Vidal, Zhi-Qiang You, Ying Zhu, Bushra Alam, Benjamin J. Albrecht, Abdulrahman Aldossary, Ethan Alguire, Josefine H. Andersen, Vishikh Athavale, Dennis Barton, Khadiza Begam, Andrew Behn, Nicole Bellonzi, Yves A. Bernard, Eric J. Berquist, Hugh G. A. Burton, Abel Carreras, Kevin Carter-Fenk, Romit Chakraborty, Alan D. Chien, Kristina D. Closser, Vale Cofer-Shabica, Saswata Dasgupta, Marc de Wergifosse, Jia Deng, Michael Diedenhofen, Hainam Do, Sebastian Ehlert, Po-Tung Fang, Shervin Fatehi, Qingguo Feng, Triet Friedhoff, James Gayvert, Qinghui Ge, Gergely Gidofalvi, Matthew Goldey, Joe Gomes, Cristina E. González-Espinoza, Sahil Gulania, Anastasia O. Gunina, Magnus W. D. Hanson-Heine, Phillip H. P. Harbach, Andreas Hauser, Michael F. Herbst, Mario Hernández Vera, Manuel Hodecker, Zachary C. Holden, Shannon Houck, Xunkun Huang, Kerwin Hui, Bang C. Huynh, Maxim Ivanov, Adam Jasz, Hyunjun Ji, Hanjie Jiang, Benjamin Kaduk, Sven Kähler, Kirill Khistyaev, Jaehoon Kim, Gergely Kis, Phil Klunzinger, Zsuzsanna Koczor-Benda, Joong Hoon Koh, Dimitri Kosenkov, Laura Koulias, Tim Kowalczyk, Caroline M. Krauter, Karl Kue, Alexander Kunitsa, Thomas Kus, Istvan Ladjanszki, Arie Landau, Keith V. Lawler, Daniel Lefrancois, Susi Lehtola, Run R. Li, Yi-Pei Li, Jiashu Liang, Marcus Liebenthal, Hung-Hsuan Lin, You-Sheng Lin, Fenglai Liu, Kuan-Yu Liu, Matthias Loipersberger, Arne Luenser, Aaditya Manjanath, Prashant Manohar, Erum Mansoor, Sam F. Manzer, Shan-Ping Mao, Aleksandr V. Marenich, Thomas Markovich, Stephen Mason, Simon A. Maurer, Peter F. McLaughlin, Maximilian F. S. J. Menger, Jan-Michael Mewes, Stefanie A. Mewes, Pierpaolo Morgante, J. Wayne Mullinax, Katherine J. Oosterbaan, Garrette Paran, Alexander C. Paul, Suranjan K. Paul, Fabijan Pavošević, Zheng Pei, Stefan Prager, Emil I. Proynov, Adam Rak, Eloy Ramos-Cordoba, Bhaskar Rana, Alan E. Rask, Adam Rettig, Ryan M. Richard, Fazle Rob, Elliot Rossomme, Tarek Scheele, Maximilian Scheurer, Matthias Schneider, Nickolai Sergueev, Shaama M. Sharada, Wojciech Skomorowski, David W. Small, Christopher J. Stein, Yu-Chuan Su, Eric J. Sundstrom, Zhen Tao, Jonathan Thirman, Gábor J. Tornai, Takashi Tsuchimochi, Norm M. Tubman, Srimukh Prasad Veccham, Oleg Vydrov, Jan Wenzel, Jon Witte, Atsushi Yamada, Kun Yao, Sina Yeganeh, Shane R. Yost, Alexander Zech, Igor Ying Zhang, Xing Zhang, Yu Zhang, Dmitry Zuev, Alán Aspuru-Guzik, Alexis T. Bell, Nicholas A. Besley, Ksenia B. Bravaya, Bernard R. Brooks, David Casanova, Jeng-Da Chai, Sonia Coriani, Christopher J. Cramer, György Cserey, A. Eugene DePrince, Robert A. DiStasio, Andreas Dreuw, Barry D. Dunietz, Thomas R. Furlani, William A. Goddard, Sharon Hammes-Schiffer, Teresa Head-Gordon, Warren J. Hehre, Chao-Ping Hsu, Thomas-C. Jagau, Yousung Jung, Andreas Klamt, Jing Kong, Daniel S. Lambrecht, WanZhen Liang, Nicholas J. Mayhall, C. William McCurdy, Jeffrey B. Neaton, Christian Ochsenfeld, John A. Parkhill, Roberto Peverati, Vitaly A. Rassolov, Yihan Shao, Lyudmila V. Slipchenko, Tim Stauch, Ryan P. Steele, Joseph E. Subotnik, Alex J. W. Thom, Alexandre Tkatchenko, Donald G. Truhlar, Troy Van Voorhis, Tomasz A. Wesolowski, K. Birgitta Whaley, H. Lee Woodcock, Paul M. Zimmerman, Shirin Faraji, Peter M. W. Gill, Martin Head-Gordon, John M. Herbert, and Anna I. Krylov. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.*, 155(8):084801, August 2021. Publisher: American Institute of Physics.

[44] Frank Weinhold and Eric D Glendening. Nbo 5.0 program manual: natural bond orbital analysis programs. *Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin, Madison, WI*, 53706, 2001.

[45] Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal Neighbourhood Aggregation for Graph Nets. *arXiv:2004.05718 [cs, stat]*, December 2020. arXiv: 2004.05718.

[46] Guohao Li, Matthias Müller, Ali Thabet, and Bernard Ghanem. DeepGCNs: Can GCNs Go as Deep as CNNs? *arXiv:1904.03751 [cs]*, August 2019. arXiv: 1904.03751.

[47] Jonathan Godwin, Michael Schaarschmidt, Alexander Gaunt, Alvaro Sanchez-Gonzalez, Yulia Rubanova, Petar Veličković, James Kirkpatrick, and Peter Battaglia. Simple GNN Regularisation for 3D Molecular Property Prediction & Beyond. *arXiv:2106.07971 [cs]*, March 2022. arXiv: 2106.07971.

[48] Chen Cai and Yusu Wang. A Note on Over-Smoothing for Graph Neural Networks. *arXiv:2006.13318 [cs, stat]*, June 2020. arXiv: 2006.13318.