JCTC Journal of Chemical Theory and Computation

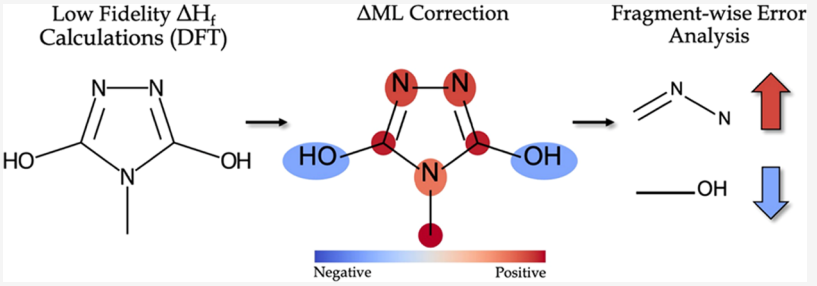
pubs.acs.org/JCTC Article

Interpretable Graph-Network-Based Machine Learning Models via Molecular Fragmentation

Eric M. Collins* and Krishnan Raghavachari*

Cite This: https://doi.org/10.1021/acs.jctc.2c01308 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Chemists have long benefitted from the ability to understand and interpret the predictions of computational models. With the current shift to more complex deep learning models, in many situations that utility is lost. In this work, we expand on our previously work on computational thermochemistry and propose an interpretable graph network, FragGraph(nodes), that provides decomposed predictions into fragment-wise contributions. We demonstrate the usefulness of our model in predicting a correction to density functional theory (DFT)-calculated atomization energies using Δ-learning. Our model predicts G4(MP2)-quality thermochemistry with an accuracy of <1 kJ mol^{-1} for the GDB9 dataset. Besides the high accuracy of our predictions, we observe trends in the fragment corrections which quantitatively describe the deficiencies of B3LYP. Node-wise predictions significantly outperform our previous model predictions from a global state vector. This effect is most pronounced as we explore the generality by predicting on more diverse test sets indicating node-wise predictions are less sensitive to extending machine learning models to larger molecules.

1. INTRODUCTION

Machine learning (ML) models have become increasingly more available and used for a wide variety of applications in diverse disciplines.[1−15] Although these models have become invaluable for many, the field is moving towards deeper, more complex models. While this is advantageous to learn complex patterns, oftentimes the models are black-box and their predictions cannot easily be understood or explained by the user.[16−21] In contemporary applications, complex models are necessary for the advanced tasks on which machine learning and artificial intelligence are being applied. While acceptable in data-driven applications such as computer vision and language models, scientific applications would highly benefit from a more explainable set of predictions. Furthermore, the adoption of some state of the art graph-based models has been slow for certain applications such as drug discovery, in part, for this lack of interpretability.[21] Chemists and other scientists have used simple statistical models, such as regression and decision trees, long before the rise in popularity of deep learning models. These models and algorithms have an inbuilt explainability and interpretability due to their overall simplicity. These attributes have allowed scientists to reason through many problems and form a more advanced chemical intuition.

In our published work in 2021, we developed a graph-network based deep learning model, FragGraph. This model was used in a manner similar to a fingerprint encoder in which a global (state) vector was learned from the structure and features of a molecular graph. This work showed the benefit from embedding node attributes to represent local fragments rather than starting from a simple graph and allowing the model to learn its own representation. Although we showed its excellent performance at predicting G4(MP2)-quality thermochemistry, the model provided little insight into the learned patterns in the training data. In the graph neural network model used in FragGraph, updated edge, node, and global state vectors are all learned through each graph update step. After the final graph update, a full graph is embedded into its latent space representation. Although this can be advantageous for generality, since multiple decoders could be used to predict any edge-, node-, or state-wise

Received: December 24, 2022

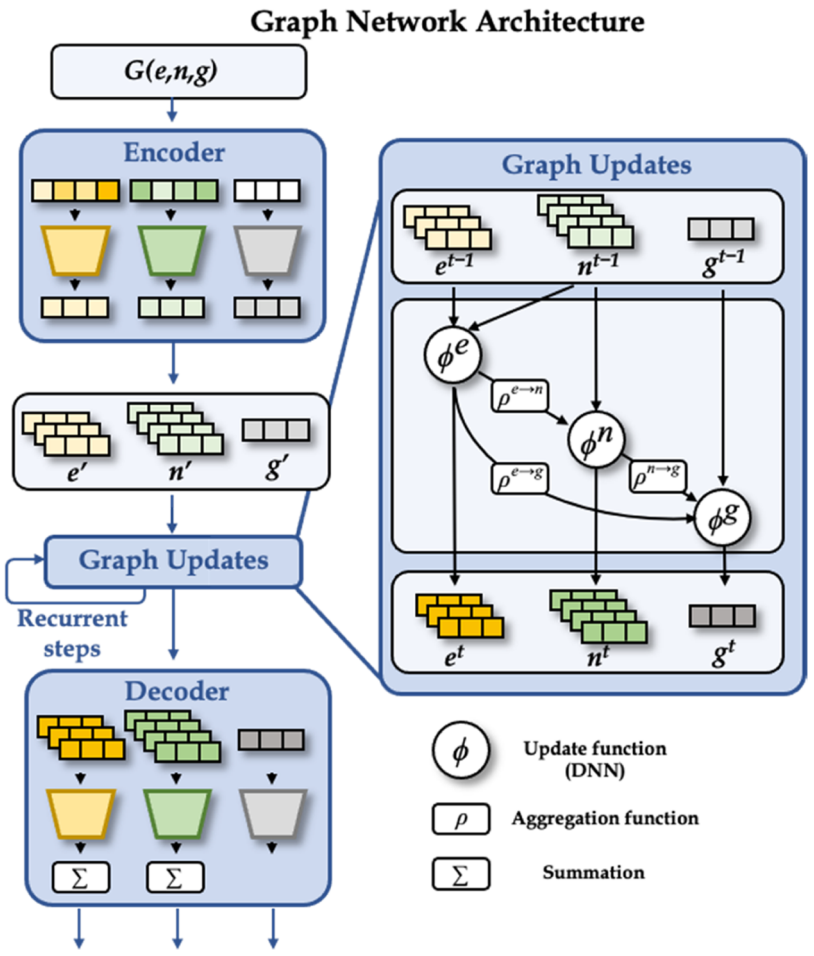


Figure 1. Architecture schematic for graph network.

property, the full extent of the graph network's predictability has yet to be explored. In this work, we explore and extend the usefulness of the FragGraph model by predicting the machine learned corrections to DFT as a sum of node-wise contributions.

There is a growing number of research groups working towards developing explainable artificial intelligence (XAI) techniques which can be applied to scientific problems. In drug discovery, for example, the integrated gradients approach was used in conjunction with a graph network to identify pharmacophore motifs and activity cliffs by assigning importance to structural features in the predicted pharmacology-related end points.[22] Some graph-based methods aim to identify node importance through various down sampling or pooling layers and have been used to distinguish important structural features in relation to properties such as molecular toxicity.[23−25] Other methods have been used to identify important molecular descriptors for the prediction of the permeabilities of polymer membranes using the SHAP analysis.[26,27]

While these methodologies have seen success in providing a qualitative interpretation, this work aims to reframe the interpretability to obtain an explicit quantitative picture along with the useful qualitative explanation of the molecular predictions. Herein, we strive to develop on our recent FragGraph model into an explainable machine learning model in which a prediction can be decomposed into contributions based on chemical structure. We focus on correcting the errors of approximate methods, i.e., DFT, and draw parallels to fragmentation-based methodology. The fundamental nature of fragments as they relate to systematic error correction allows for the identification of deficiencies in the chosen approximate method and we propose our model may be further utilized to study the deficiencies of other approximate methods.

2. METHODS

The present study focuses on two FragGraph models. To minimize the number of variables between the two models, all latent space sizes and neural network layers are kept consistent throughout each architecture. The general architecture used for this work is similar to the previously used FragGraph model consisting of seven individual neural networks, with the difference being solely in the decoder stage. An overview of the architecture is shown in Figure 1 and in more detail in our previous work. The model consists of three parts: encoder, graph update, and decoder. In each portion, there are up to three neural networks corresponding to the three parts of a molecular graph. The encoder converts the initial representation into the correct latent space. Then, message-passing graph updates are performed for each node, edge, and global vector (in that order) to learn from the local graph structure and attributes. After 3 recurrent steps, the graph is finalized and can be used to predict any property of the system via a readout in the decoder stage.

The first model used in this work is identical to the global FragGraph model in our previous work, while the second model incorporates node-wise predictions for the atomization energy using a shared decoder neural network for all atom types.

https://doi.org/10.1021/acs.jctc.2c01308
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

For the FragGraph representation, each heavy atom is represented as a node in a molecular graph, and bonds between them are represented as edges. These nodes are then attributed with a local description of the chemical environment. In the previous work, we demonstrated the importance of the initial description of the chemical environment. The fragment embedded graph performed with an accuracy of around 0.5 kJ mol^{-1} for the prediction of atomization energy and a simple graph with atomic number information was only able to achieve around 2 kJ mol^{-1} on average. Although impressive, the automatic feature-learning present in many deep learning models benefitted greatly from information about the local fragment. Thus, atomic environment representations of CBH-2 fragments were used by passing each fragment through the pretrained mol2vec model. The molecular graphs used here are complete graphs, i.e., each node is connected to every other node. These fully connected graphs provide more three-dimensional (3D) structural information to the model and allow it to learn from non-bonded interactions as well as bonded. Edges are encoded with one-hot encoded atom types involved in the bond, bond order (non-bonded 0, single bond 1, double bond 2, or triple bond 3), as well as gaussian expanded bond distance between atom centers. Although the new models have the capability to return edge and global feature vectors, since there are no direct decoders, they are disregarded. Additionally, since the global vector is not involved in the update of the nodes and edges, the new FragGraph models do not have a global vector.

The main dataset used for this work is the openly available GDB9 set of 130k molecules containing up to 9 C, N, O, and F heavy atoms.[6,28] This dataset was curated to represent the chemical space of small organic molecules including small amino acids and pharmaceutically relevant building blocks. All 130K molecules are neutral species with 1705 of them being zwitterionic. The models featured herein were trained on 117k training molecules with the remaining 13k of the GDB9 dataset acting as the out-of-sample generalization set. The same train-test split was used from previous studies for consistency.[29,30] Two external test sets were also used to test the generality of each model. These include the GDB10−13 data set of 1500 molecules with 10−13 heavy atoms sampled from the GDB17 set.[31] Additionally, the PDS10−14 set of 191 molecules with 10−14 non-hydrogen atoms was also used.[32] This test set was curated from the Pedley compilation of experimental gas-phase enthalpy of formation data for organic molecules. The 191 molecules were chosen as having experimental values with low uncertainty 1 kcal mol^{-1} as well as a close agreement with the G4(MP2) calculated enthalpies. All molecules were optimized with B3LYP/6-31G(2df,p), and the atomization energies of all molecules were calculated with both G4(MP2) and B3LYP.[33,34]

3. RESULTS AND DISCUSSION

3.1. FragGraph Model Performance. The performance of the two FragGraph models is summarized in Table 1. Each model had the same training set of 117k molecules from the GDB9 dataset and all displayed numbers are mean absolute errors of holdout sets. On the GDB9 test set, the new FragGraph(nodes) model slightly outperforms the original global model with a mean absolute error (MAE) of 0.16 kcal mol^{-1} compared to 0.18 kcal mol^{-1}. These results are consistent to the previous FragGraph study.[30] Both models had errors close to zero. Additionally, we found that no further benefit is gained from adding decoder complexity by having a different neural network for each atom type. Thus, moving forward the decoder will be kept as the single shared neural network for all atom types.

Table 1. Mean Absolute Errors in kcal mol^{-1} of FragGraph(global) and FragGraph(nodes) on the GDB9, GDB10−13, and PDS10−14 Datasets

model	GDB9 (N = 13 024)	GDB10−13 (N = 1500)	PDS10−14 (N = 191)
FragGraph(global)	0.18	1.19	1.25
FragGraph(nodes)	0.16	1.01	1.07

While the performance is excellent on small molecules, the errors increase substantially for predicting the atomization energies of larger molecules with about a 7-fold increase in errors. We note that a slight increase in error is typically expected for larger molecules, as the systematic errors will grow with system size. Indeed, scaling each error by the number of heavy atoms results in a MAE of 0.020 kcal mol^{-1} $atom^{-1}$ for the GDB9 data set along with 0.087 and 0.094 kcal mol^{-1} $atom^{-1}$ for the GDB10−13 and PDS10−14 datasets, respectively. Although there is still a discrepancy between the $N_{HA} < 10$ and $N_{HA} > 9$ groups, this difference is lowered to ∼4-fold increase in error rather than the previous 7-fold.

Machine learning models tend to perform best on interpolation rather than extrapolation. In principle, one could train a general model which would work on any system regardless of size or composition to rival modern quantum mechanical methods, but this is currently unfeasible. Additionally, the FragGraph models in this case may be overfit to the GDB9 dataset and are not size extensive since training on small molecules and predicting on larger molecules is more of an extrapolation. To test this hypothesis, larger molecules from the GDB10−13 test set were added to the training set. We tested adding various percentages including 20, 50, and 80%. Even at the smallest addition to the training set, the test error on the GDB10−13 data set dropped from 1.01 to 0.67 kcal mol^{-1} for the FragGraph(nodes) model. This decrease is somewhat surprising since the ratio of larger molecules to the $N_{HA} < 10$ is about 1:400. Upon training with 50% of the larger molecule test set, the performance improved slightly to a MAE 0.56 kcal mol^{-1}. Adding additional molecules to the training set showed little to no improvement leading to the final FragGraph models in Table 2. The performance on the completely held out PDS10−14 test set also improved by 0.4 kcal mol^{-1} for the global model and 0.3 kcal mol^{-1} for the node-wise FragGraph model.

Table 2. Mean Absolute Errors in kcal mol^{-1} of FragGraph(global) and FragGraph(nodes) on the GDB9, GDB10−13, and PDS10−14 Datasets with Larger Molecules Added to the Training Set

model	GDB9 (N = 13 024)	GDB10−13 (N = 300)	PDS10−14 (N = 191)
FragGraph(global)	0.18	0.58	0.87
FragGraph(nodes)	0.17	0.53	0.76

FragGraph models trained solely on the $N_{HA} < 10$ data set performed similarly for both the GDB10−13 and PDS10−14 test sets. However, this is not case upon the addition of larger molecules in the training set. Since we have already accounted for poor performance due to size, there must be another source of error. To further understand where the errors from the

https://doi.org/10.1021/acs.jctc.2c01308
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

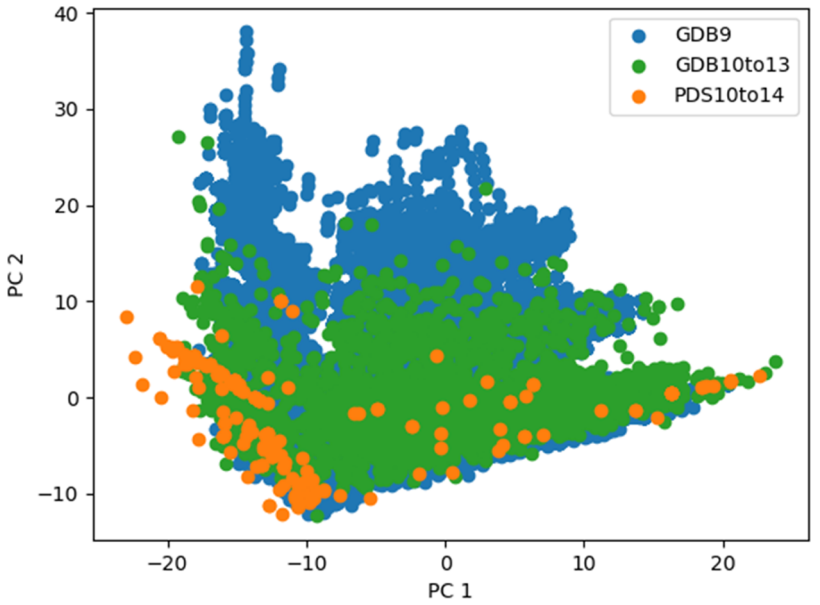


Figure 2. Principal component analysis distribution of the GDB9, GDB10−13, and PDS10−14 datasets.

PDS10−14 dataset are coming from, we can visualize the datasets using principal component analysis (PCA). The global mol2vec vector was generated for each molecule and reduced to 2 principal components, shown in Figure 2.

Unsurprisingly, the GDB9 and GDB10−13 datasets had roughly similar distributions, most likely because they were both sampled from the same parent collection. While the majority of the distribution of the PDS10−14 test set falls within the other two distributions, there are many more outlying molecules. Upon further inspection, these systems contain fragments which were previously unseen by the model. For example, one of the many molecules below PC 1 = −20, CCCCCCC(N(F)F)N-(F)F, contains two fluorinated nitrogen groups. This shows the GDB9 training is insufficient for generality, or at least in the case of the PDS10−14 test set.

We observed that if we sample the PDS10−14 to include these unseen fragments, the performance of the FragGraph model increased further to 0.57 kcal mol^{-1} on average for the node-wise model, which agrees more with the GDB10−13 dataset. The performance of these models is shown in Table 3

Table 3. Mean Absolute Errors in kcal mol^{-1} of FragGraph(global) and FragGraph(nodes) on the GDB9, GDB10−13, and PDS10−14 Datasets with Larger and More Diverse Molecules Added to the Training Set

model	GDB9 (N = 13 024)	GDB10−13 (N = 300)	PDS10−14 (N = 171)
FragGraph(global)	0.18	0.56	0.75
FragGraph(nodes)	0.17	0.52	0.57

with only 10% of the PDS10−14 set added to the training set. These molecules were sampled as the furthest distance (or most dissimilar) molecules from the GDB9 distribution. The global model still had a larger discrepancy, which may be due to the inability for one global vector to capture newer fragments. We propose the node-wise model is more flexible for generality than the global model due to the node-wise contributions.

3.2. Model Explainability. One unique advantage to the node-wise predictions is the added explainability. Black-box machine learning models are often criticized for the inability to understand why they are making each prediction. Although this mystery is not necessarily a downside for many models, having the ability to explain and understand a prediction can be a plus. The FragGraph(nodes) model inherently provides explainability since it returns the node-wise contributions for any given molecule. Since the models are trained on the difference between B3LY and G4(MP2), these numbers correlate to an atom-wise or fragment-wise correction to DFT. For example, the node-wise contributions are given for one of the molecules from the GDB9 test set in Figure 3. The enthalpy of formation for this

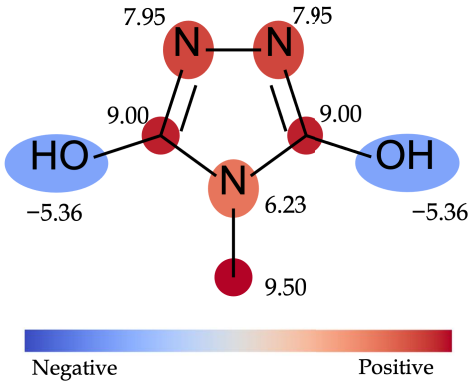


Figure 3. Fragment-wise contributions from the trained FragGraph model for a triazole compound from the GDB9 test set.

molecule (calculated at G4(MP2)) is −139.20 kJ mol^{-1}, while the DFT calculated value is −177.86 kJ mol^{-1} giving an approximate error of 38 kJ mol^{-1}. Using the FragGraph model, the sum of these contributions make a ΔML correction of 39 kJ mol^{-1}, predicting the enthalpy within 1 kJ mol^{-1} of the reference value.

These generated heat maps give a visual representation of the error correction and can show which groups contribute most to the approximation errors of DFT. Since the FragGraph model is rooted in fragmentation and error cancellation, these contributions can be directly correlated with the initial fragment each

https://doi.org/10.1021/acs.jctc.2c01308
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

node represents. In the previous example, our FragGraph model learned a large positive correction for N-containing heterocycles and a large negative contribution for hydroxyl groups.

Traditionally in fragmentation-based methods, each fragment contributes a set value calculated as the difference between two levels of theory. Our FragGraph model is not restricted to a single value per fragment, since no explicit energy calculations are performed on the fragments. For any given initial fragment, a distribution of contributions will be learned based on the surrounding fragments. In this way, the graph network learns from somewhat of a larger fragment space and can augment the correction based on the relationships between fragments in the same molecule. For example, the common propene-like fragment, shown in Figure 4, would be restricted to a single

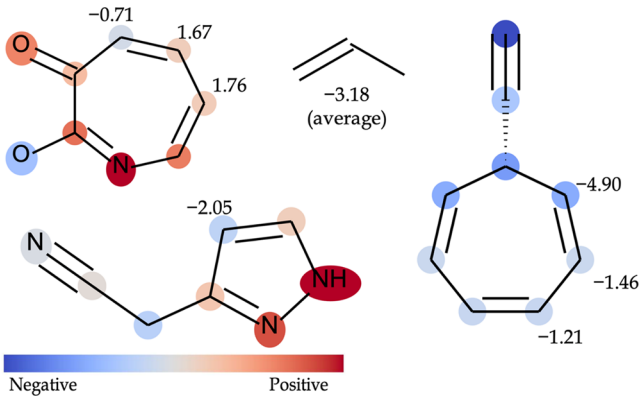


Figure 4. Fragment-wise contributions from the trained FragGraph model for a variety of propene-like fragment containing molecules.

value in the traditional fragmentation sense. As seen in three molecules displayed in Figure 4, this fragment contribution value can have a large distribution for more negatively contributing groups such as the nitrile group or contribute more positively when found in a N-containing heterocycle (Figure 4).

The full fragment-wise contribution statistics are given in Supporting Information. For each initial fragment formed from a heavy atom, the average and standard deviation of the final node-wise contributions were calculated for the full GDB9 dataset. Interestingly, patterns in the fragments contributions and composition begin to emerge. This is illustrated in Table 4a

Table 4a. Average Fragment Contributions from the FragGraph(nodes) Model for Alkyne and Some Branched Fragments

fragment	contribution
C≡C	−9.92
CC(C)(O)O	−5.52
CC(C)C	−3.26
CC(C)(C)O	−3.18
CC(C)(C)N	−3.02

for a select set of illustrative fragments. For example, branched systems and fragments containing triple bonds tend to contribute more negatively while more highly energetic functional groups such as fragments containing multiple nitrogen atoms or fluorine tend to contribute more positively to the full correction. Since the models in the study were trained on the difference between B3LYP and G4(MP2), these fragment-wise contributions shed light into the errors of B3LYP. Indeed, many studies have pointed out the deficiencies of the popular density functional B3LYP including systematically underestimating the heat of formation of hydrocarbons.[35−37] Furthermore, these errors are even more pronounced in larger, branched systems. B3LYP and several density functionals often fail at correctly capturing medium-range electron correlation and many ongoing developments in DFT have focused on correcting for this well-known error.[38,39] Our results here are in line with this observation, as many of the branched fragments correct for this underestimation. On the other end of the spectrum, B3LYP over-stabilizes the heat of formation for highly energetic systems leading to a positive correction to the thermochemical properties (Table 4b).

Table 4b. Average Fragment Contributions from the FragGraph(nodes) Model for Some Highly Energetic Fragments

fragment	contribution
CC(F)(F)F	19.97
CN(N)N	18.31
CN═O	18.15
NC(N)N	16.90
CN(C)N	16.86

4. CONCLUSIONS

The present work illustrates the utility of our interpretable fragment-based graph machine learning model FragGraph. Using the sum of contributions from node-based vectors was found to be more effective at generalizing to large and more complex molecules than the corresponding global vector prediction model. Our model achieves excellent results on molecules with less than 10 heavy atoms at around 0.17 kcal mol^{-1} compared to high level reference heats of formation and approximately 0.5 kcal mol^{-1} on molecules larger than 10 heavy atoms and a MAE of 0.6 kcal mol^{-1} on the challenging Pedley data set. Through training set design and unseen fragment sampling, we were able to improve the predictive value to an error of 0.04 kcal mol^{-1} per heavy atom. Additionally, our model was able to replicate known trends in the deficiencies of B3LYP. We propose the FragGraph model can be further utilized for other density functionals to understand the systematic errors in a structure-based manner and could be useful in developing new models.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c01308.

Fragment contribution statistics for the FragGraph-(nodes) model (Table S1) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Eric M. Collins − *Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States;* orcid.org/0000-0002-9113-1705; Email: colliner@iu.edu

Krishnan Raghavachari − *Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States;*

https://doi.org/10.1021/acs.jctc.2c01308
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

orcid.org/0000-0003-3275-1426; Email: kraghava@indiana.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c01308

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge financial support from the National Science Foundation Grant CHE-2102583 at Indiana University. The Big Red 3 supercomputing facility at Indiana University was used for most of the calculations in this study.

■ REFERENCES

(1) Gori, M.; Monfardini, G.; Scarselli, F. In *A New Model for Learning in Graph Domains*, Proceedings 2005 IEEE International Joint Conference on Neural Networks, 2005; Vol. 7922, pp 729−734.

(2) Rupp, M.; Tkatchenko, A.; Muller, K. R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, No. 058301.

(3) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, No. 140022.

(4) Perozzi, B.; Al-Rfou, R.; Skiena, S. In *DeepWalk: Online Learning of Social Representations*, Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Association for Computing Machinery: New York: New York, USA, 2014; pp 701−710.

(5) Jordan, M. I.; Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255−260.

(6) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Delta-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−2096.

(7) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Muller, K. R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326−2331.

(8) Ekins, S. The Next Era: Deep Learning in Pharmaceutical Research. *Pharm. Res.* **2016**, *33*, 2594−2603.

(9) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *Npj Comput Mater* **2017**, *3*, No. 54.

(10) Li, Y.; Yu, R.; Shahabi, C.; Liu, Y.Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting, 2017. arXiv:1707.01926. arXiv.org e-Print archive. https://doi.org/10.48550/arXiv.1707.01926.

(11) Lo, Y. C.; Rensi, S. E.; Torng, W.; Altman, R. B. Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today* **2018**, *23*, 1538−1546.

(12) Battaglia, P. W.; Hamrick, J. B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R.; Gulcehre, C.; Song, F.; Ballard, A.; Gilmer, J.; Dahl, G.; Vaswani, A.; Allen, K.; Nash, C.; Langston, V.; Dyer, C.; Heess, N.; Wierstra, D.; Kohli, P.; Botvinick, M.; Vinyals, O.; Li, Y.; Pascanu, R. Relational Inductive Biases, Deep Learning, and Graph Networks, 2018. arXiv:1806.01261. arXiv.org e-Print archive. https://doi.org/10.48550/arXiv.1707.01926.

(13) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discovery* **2019**, *18*, 463−477.

(14) Harada, S.; Akita, H.; Tsubaki, M.; Baba, Y.; Takigawa, I.; Yamanishi, Y.; Kashima, H. Dual graph convolutional neural network for predicting chemical networks. *BMC Bioinf.* **2020**, *21*, No. 94.

(15) von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **2020**, *4*, 347−358.

(16) Roscher, R.; Bohn, B.; Duarte, M. F.; Garcke, J. Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* **2020**, *8*, 42200−42216.

(17) Belle, V.; Papantonis, I. Principles and Practice of Explainable Machine Learning. *Front. Big Data* **2021**, *4*, No. 39.

(18) Samek, W.; Wiegand, T.; Müller, K.-R. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models, 2017. arXiv:1708.08296. arXiv.org e-Print archive. https://doi.org/10.48550/arXiv.1708.08296.

(19) von Rueden, L.; Mayer, S.; Beckh, K.; Georgiev, B.; Giesselbach, S.; Heese, R.; Kirsch, B.; Pfrommer, J.; Pick, A.; Ramamurthy, R.; Walczak, M.; Garcke, J.; Bauckhage, C.; Schuecker, J. Informed Machine Learning – A Taxonomy and Survey of Integrating Knowledge into Learning Systems. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 614−633.

(20) Doshi-Velez, F.; Kim, B.; Towards, A.Towards a Rigorous Science of Interpretable Machine Learning, 2017. arXiv:1702.08608. arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608.

(21) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573−584.

(22) Jiménez-Luna, J.; Skalic, M.; Weskamp, N.; Schneider, G. Coloring Molecules with Explainable Artificial Intelligence for Preclinical Relevance Assessment. *J. Chem. Inf. Model.* **2021**, *61*, 1083−1094.

(23) Noutahi, E.; Beaini, D.; Horwood, J.; Giguère, S.; Tossou, P. Towards Interpretable Sparse Graph Representation Learning with Laplacian Pooling, 2019. arXiv:1905.11577. arXiv.org e-Print archive. https://doi.org/10.48550/arXiv.1905.11577.

(24) Ying, R.; You, J.; Morris, C.; Ren, X.; Hamilton, W. L.; Leskovec, J. Hierarchical Graph Representation Learning with Differentiable Pooling. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2018; Vol. *31*.

(25) Gao, H.; Ji, S. In *Graph U-Nets*, International Conference on Machine Learning; PMLR, 2019; pp 2083−2092.

(26) Yang, J.; Tao, L.; He, J.; McCutcheon, J. R.; Li, Y. Machine learning enables interpretable discovery of innovative polymers for gas separation membranes. *Sci. Adv.* **2022**, *8*, No. eabn9545.

(27) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2017; Vol. *30*.

(28) Narayanan, B.; Redfern, P. C.; Assary, R. S.; Curtiss, L. A. Accurate quantum chemical energies for 133 000 organic molecules. *Chem. Sci.* **2019**, *10*, 7449−7455.

(29) Ward, L.; Blaiszik, B.; Foster, I.; Assary, R. S.; Narayanan, B.; Curtiss, L. Machine learning prediction of accurate atomization energies of organic molecules from low-fidelity quantum chemical calculations. *MRS Commun.* **2019**, *9*, 891−899.

(30) Collins, E. M.; Raghavachari, K. A Fragmentation-Based Graph Embedding Framework for QM/ML. *J. Phys. Chem. A* **2021**, *125*, 6872−6880.

(31) 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. *J. Chem. Inf. Model.* **2012**, *52*, 2864−2875.

(32) Dandu, N.; Ward, L.; Assary, R. S.; Redfern, P. C.; Narayanan, B.; Foster, I. T.; Curtiss, L. A. Quantum-Chemically Informed Machine Learning: Prediction of Energies of Organic Molecules with 10 to 14 Non-hydrogen Atoms. *J. Phys. Chem. A* **2020**, *124*, 5804−5811.

(33) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory using reduced order perturbation theory. *J. Chem. Phys.* **2007**, *127*, No. 124105.

(34) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(35) Check, C. E.; Gilbert, T. M. Progressive Systematic Underestimation of Reaction Energies by the B3LYP Model as the Number of C−C Bonds Increases: Why Organic Chemists Should Use Multiple

DFT Models for Calculations Involving Polycarbon Hydrocarbons. *J. Org. Chem.* **2005**, *70*, 9828−9834.
(36) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.* **1997**, *106*, 1063−1079.
(37) Redfern, P. C.; Zapol, P.; Curtiss, L. A.; Raghavachari, K. Assessment of Gaussian-3 and Density Functional Theories for Enthalpies of Formation of C1−C16 Alkanes. *J. Phys. Chem. A* **2000**, *104*, 5850−5854.
(38) Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157−167.
(39) Grimme, S. Seemingly Simple Stereoelectronic Effects in Alkane Isomers and the Implications for Kohn−Sham Density Functional Theory. *Angew. Chem., Int. Ed.* **2006**, *45*, 4460−4464.

https://doi.org/10.1021/acs.jctc.2c01308
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX