Gated Graph Recursive Neural Networks for Molecular Property Prediction

Hiroyuki Shindo
Division of Information Science
Nara Institute of Science and Technology
shindo@is.naist.jp

Yuji Matsumoto
Division of Information Science
Nara Institute of Science and Technology
yuji@is.naist.jp

Abstract

Molecule property prediction is a fundamental problem for computer-aided drug discovery and materials science. Quantum-chemical simulations such as density functional theory (DFT) have been widely used for calculating the molecule properties, however, because of the heavy computational cost, it is difficult to search a huge number of potential chemical compounds. Machine learning methods for molecular modeling are attractive alternatives, however, the development of expressive, accurate, and scalable graph neural networks for learning molecular representations is still challenging. In this work, we propose a simple and powerful graph neural networks for molecular property prediction. We model a molecular as a directed complete graph in which each atom has a spatial position, and introduce a recursive neural network with simple gating function. We also feed input embeddings for every layers as skip connections to accelerate the training. Experimental results show that our model achieves the state-of-the-art performance on the standard benchmark dataset for molecular property prediction.

1 Introduction

Predicting the properties of molecules is a crucial ingredient for computer-aided drug discovery and materials development with desired properties [8, 13, 16, 29]. Currently, quantum-chemical simulations based on density functional theory (DFT) [2, 5] are widely used to calculate the electronic structure and properties of molecules. However, because of the heavy computational cost of DFT, they are difficult to extensively explore the huge number of potential chemical compounds [13]. To enlarge the search space, much effort has been made to apply machine learning techniques for learning molecular representations in cheminformatics and materials informatics, while it has not been fully developed. Efficient and accurate machine learning methods for the prediction of molecular properties can have a huge impact on the discovery of novel drugs and materials.

Previous work on molecular modeling with machine learning methods has mainly focused on developing hand-crafted features for molecular representations that can reflect structural similarities and biological activities of molecules. Examples include Extended-Connectivity Fingerprints [25], Coulomb Matrix [26], Symmetry Function [3], and Bag-of-Bonds [15]. These molecular representations can be used for the prediction of molecular properties with logistic regression and kernel methods.

Recently, deep learning methods have gained a lot of attention for learning molecular representations, with the availability of large scale training data generated by quantum-chemical simulations [13]. In particular, graph neural networks are reasonable and attractive approaches since they can learn appropriate molecular representations that are invariant to graph isomorphism in an end-to-end fashion. While a number of graph neural networks have been proposed and applied to molecular modeling,

developing accurate and scalable neural networks enough to express a variety of molecules is still a challenging problem.

In this work, we present a simple and powerful graph neural network, gated graph recursive neural networks (GGRNet), for learning molecular representations and predicting molecular properties. To construct an expressive and accurate neural network, we model a molecule as a complete directed graph in which each atom has a three-dimensional coordinates, and update hidden vectors of atoms depending on the distances between them. In our model, the parameters for learning hidden atom vectors are shared across all layers, and the input embeddings are fed into every layers as skip connections to accelerate the training. Our model also allows to incorporate arbitrary features such as the number of atoms in the molecule, which is helpful for learning better representations of molecules.

We validate our model on three benchmark datasets for molecular property prediction: QM7b, QM8, and QM9 and empirically show that our model achieves superior performance than conventional methods, which highlights the potential of our model for molecular graph learning.

2   Related Work

Molecular fingerprints   In cheminformatics, hand-crafted features for molecules, referred to as molecular fingerprints, have been actively developed for encoding the structure of molecules [3, 6, 15, 25, 26]. These molecular fingerprints are typically a binary or integer vector that represents the presence of particular substructures in the molecule, and can be used as feature vectors for the prediction of molecular properties with machine learning [11, 19].

For example, in Extended-Connectivity Circular Fingerprints (ECFP) [25], atoms are initially assigned to integer identifiers, then the identifiers are iteratively updated with neighboring atoms and collected into the fingerprint set. Bag-of-bonds descriptor [15], inspired by "bag-of-words" featurization in natural language processing, collects "bags" that correspond to different types of bonds such as "C-C" and "C-H", and each bond in the bag is vectorized as $Z_i Z_j / |\mathbf{R}_i - \mathbf{R}_j|$ where Z_i and Z_j are the nuclear charges, while $\mathbf{R}_i$ and $\mathbf{R}_j$ are the positions of the two atoms in the bond.

Duvenaud et al. [10] generalized the computation of conventional fingerprints to be differentiable and learnable via backpropagation. They showed that the learnable fingerprints improve the predictive accuracy of molecules compared with the traditional fingerprints.

Graph neural networks for molecules   Recently, graph neural networks have attracted a lot of attention for a wide variety of tasks, including graph link prediction [33], chemistry and biology [12, 13, 28], natural language processing [21, 30], and computer vision [7, 20]. The neural network on graphs was early proposed by Gori et al. [14] and Scarselli et al. [27], and a large number of architectures have been proposed until now.

Gilmer et al. [13] proposed neural message passing networks (MPNNs) framework for learning molecular representations and showed that many graph neural networks such as Gated Graph Neural Networks (GG-NN) [17], Interaction Networks [1], and Deep Tensor Neural Networks (DTNN) [28] fall under the framework. They tested the MPNN on QM9 dataset for the prediction of molecular properties and achieved the state-of-the-art results.

Schütt et al. [29] introduced SchNet with the continuous-filter convolutions that map the atom positions in the molecule to the corresponding filter values. The learned filters are interacted with atom features to generate more sophisticated atom representations. The continuous-filter convolutions are incorporated into graph neural networks for the prediction of molecular energy and interatomic forces. SchNet is similar to ours in that it assumes the atoms have spartial information and learns the interactions between atoms depending on the distances.

Veličković et al. [31] introduced an attention mechanism to graph neural networks. The graph attention networks aggregate the hidden representations of vertices in the graph by weighting over its neighbors, following the self-attention mechanism that is widely used in many sequential models. They show that the attention mechanism is helpful for node classification tasks on citation network and protein-protein interaction.
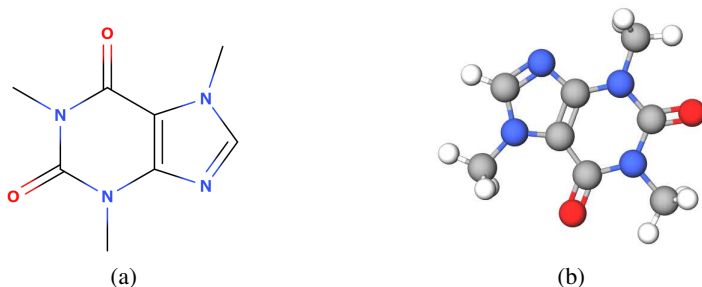
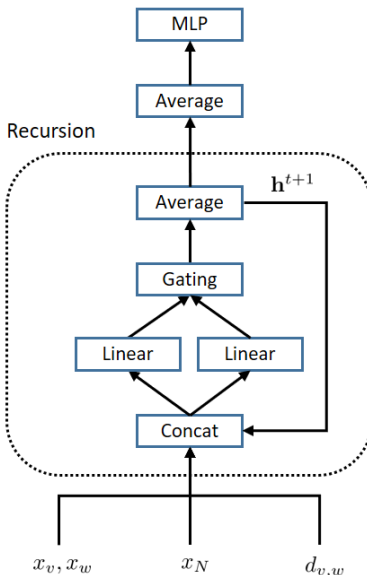Figure 1: (a) 2d structure of a molecule. (b) 3d structure of a molecule.

Figure 2: Network architecture of GGRNet

3 Gated Graph Recursive Neural Networks (GGRNet)

A molecular graph consists of atoms and chemical bonds where atoms correspond to vertices and chemical bonds correspond to edges. Figure 1 shows an example of 2D and 3D structures of a molecule. The conventional graph neural networks mainly handle with 2D structure while we treat the molecule as 3D graph structure where three-dimensional coordinates of each atom are given.

3.1 Model

In this section, we propose a gated graph recursive neural networks (GGRNet) for accurately learning molecular representations. Since our model can be formulated as a MPNN framework [13], we basically follow the notation and definitions given in [13].

To build an expressive graph neural network for molecules, one needs to stack multiple layers for learning hidden representations of atoms and bonds. However, as the number of parameters increases, it suffers from inefficient training and lower performance. Our GGRNet alleviates the problem by the following ideas: 1) the parameters for updating hidden representations are shared across all layers, 2) the input representations (atom embeddings and additional input features) are fed into every layers as skip connections to accelerate the training, and 3) the feature vector at each atom is updated by using that of every other atoms in the graph depending on the distances between atoms.

Formally, suppose we are given a molecular graph G in which each atom has a three-dimensional position. Let $\mathbf{x}_v$ and $\mathbf{x}_w$ be d-dimensional atom embeddings of a vertex v and w, respectively. In

3

our model, the hidden vector of vertex v at time t, denotes as $\mathbf{h}_v^t$, is given by the message function M and the update function U as follows:

$$\mathbf{m}_v^{t+1} = \sum_{w \neq v} M\left(\mathbf{x}_v, \mathbf{h}_v^t, \mathbf{x}_w, \mathbf{h}_w^t\right)$$

$$\mathbf{h}_v^{t+1} = U\left(\mathbf{h}_v^t, \mathbf{m}_v^{t+1}\right)$$

where the values of M are summed over all vertices w except v, which means we assume every pair of vertices in the graph has an edge and communicate each other for every time-step.

In the original MPNN, the message function M is defined as $M_t\left(\mathbf{h}_v^t, \mathbf{h}_w^t\right)$, and w as the neighbors of v. Our model extends it to always feed the input vectors $\mathbf{x}_v$ and $\mathbf{x}_w$ into the message function. Furthermore, parameters of M and U are shared across all time-steps.

After the T time-step of message updates, the readout function R aggregates all the hidden representations:

$$y = R\left(\{\mathbf{h}_v^T | v \in G\}\right)$$

where y is a target value such as molecular property.

In this work, we consider the following funcion as M:

$$\mathbf{p}_{v,w}^t = \mathbf{W}_p \cdot \left[\oplus\left(\mathbf{x}_v, \mathbf{h}_v^t, \mathbf{x}_w, \mathbf{h}_w^t, \mathbf{x}_N, \mathbf{d}_{v,w}\right)\right] + \mathbf{b}_p$$

$$\mathbf{q}_{v,w}^t = \mathbf{W}_q \cdot \left[\oplus\left(\mathbf{x}_v, \mathbf{h}_v^t, \mathbf{x}_w, \mathbf{h}_w^t, \mathbf{x}_N, \mathbf{d}_{v,w}\right)\right] + \mathbf{b}_q$$

$$\mathbf{m}_{v,w}^{t+1} = \sigma\left(\mathbf{p}_{v,w}^t\right) \odot \tanh\left(\mathbf{q}_{v,w}^t\right)$$

$$\mathbf{m}_v^{t+1} = \sum_{w \neq v} \mathbf{m}_{v,w}^{t+1}$$

where $\oplus$ is a concatenation function and σ is a sigmoid function. $\mathbf{W}_p$, $\mathbf{W}_q$, $\mathbf{b}_p$ and $\mathbf{b}_q$ are model parameters to be learned, which are shared across all time-steps. Note that we assume G as a directed graph, hence $\mathbf{m}_{v,w}^{t+1} \neq \mathbf{m}_{w,v}^{t+1}$.

In addition to $\mathbf{x}_v$, $\mathbf{h}_v^t$, $\mathbf{x}_w$, and $\mathbf{h}_w^t$, we consider two additional input features: counting feature $\mathbf{x}_N$ and distance feature $\mathbf{d}_{v,w}$. The counting feature is a real-valued embedding vector that corresponds to the number of atoms N in the graph. Therefore, the molecules with the same number of atoms share the same counting embeddings. The distance feature $\mathbf{d}_{v,w}$ is a one-dimensional vector whose value $d_{v,w}$ is the reciprocal of the Euclid distance between v and w, that is, $d_{v,w} = 1/\sqrt{(v_x - w_x)^2 - (v_y - w_y)^2 - (v_z - w_z)^2}$ where v_x, v_y, v_z are three-dimensional coodinates of v and so as w.

Initially, $\mathbf{h}_v^0 = \mathbf{0}$ for all v. The gating function: $\sigma \odot \tanh$, inspired by LSTM and gated convolutional neural networks [9], is used to extract effective features from the previous hidden vectors and input features.

The update function U is simply an average as follows:

$$\mathbf{h}_v^{t+1} = \frac{1}{N}\mathbf{m}_v^{t+1}$$

where N is the number of atoms in the molecule.

Finally, the readout function R is given as follows:

$$y = \text{MLP}\left(\frac{1}{N}\sum_{v \in G} \mathbf{h}_v^T\right)$$

4

Table 1: Hyper-parameter settings of our model for all experiments.

Hyper-parameters \ Dataset	QM7b	QM8	QM9
Atom embedding size	50	50	50
Count embedding size	50	50	50
Hidden vector size	100	100	100
Number of recursive layers T	5	5	5
Learning rate α_0	0.03	0.03	0.01
Learning rate decay k	0.01	0.01	0.05
Number of epochs	500	500	200
Batch size	10	10	10
Gradient clipping	10.0	10.0	10.0

where we average all the hidden vectors in the graph, followed by the standard three-layer MLP with ReLU activation function.

Different from the original MPNN, our model always feeds the same **x** and additional input features into the recursive time step. The impact of counting feature and distance feature is evaluated in the experimental section.

The network architecture of GGRNet is shown in figure 2.

4   Experiments

We validate the performance of our GGRNet on molecular datasets in MoleculeNet [32]. MoleculeNet is a comprehensive benchmark for molecular machine learning and it contains multiple datasets for regression and classification tasks. In this work, we use QM7b [18], QM8 [23], and QM9 [22] datasets for regression task. In QM8 and QM9, the input is a set of discrete molecular graph with spatial positions of atoms. In QM7b, only spatial positions of atoms for each molecule are available. The target is a real-valued molecule property such as the energy of the electron and the heat capacity. For details about the dataset, please refer to [32].

We compare our GGRNet against the following three baselines: graph convolutional models (GC) [32], deep tensor neural networks (DTNN) [28], and MPNN [13]. MPNN is the implementation of an edge network as message passing function and a set2set model as readout function. The edge network considers all neighbor atoms and the feature vectors of atoms are updated with gated recurrent units. In the readout phase, LSTM with attention mechanism is applied to a sequence of feature vectors to generate the final feature vector of the molecule.

These three baseline methods achieve state-of-the-art performance on a variety of tasks for molecular modeling and outperform the conventional methods using hand-crafted features. All the baseline results are taken from MoleculeNet [32]. The codes of baseline models are publicly available via DeepChem open source library [24].

The hyperparameters of our GGRNet in the experiments are shown in table 1. As shown in the table, we use almost the same hyper-parameters for every dataset. Following the MoleculeNet, we randomly split the dataset into training, validation, and test as 80/10/10 ratio. All the reported results are averaged over three independent runs.

Target properties in the training set are normalized to zero mean and unit variance using only the training set. For evaluation, the predicted values of the test set are inversely transformed using the training mean and variance. The loss function is mean squared error (MSE) between the model output and the target value. For evaluation, we use mean absolute error (MAE). We use stochastic gradient descent (SGD) for training our model. The learning rate α is given by $\alpha = \alpha_0/(1 + ke)$ where α_0, k, e are the initial learning rate, the decay rate, and the number of epochs, respectively. All the parameters including atom embeddings and counting embeddings are initialized randomly and updated during training.

5

Table 2: QM7b test set performances (MAE)

Property \ Model	Unit	DTNN [28]	GGRNet
Atomization energy (PBE0)	kcal / mol	21.5	**13.7**
Maximal absorption intensity (ZINDO) [1]	eV	1.26	**1.02**
Excitation energy at maximal absorption (ZINDO) [2]	Arbitrary	0.074	**0.072**
HOMO (ZINDO)	eV	0.192	**0.140**
LUMO (ZINDO)	eV	0.159	**0.0915**
First excitation energy (ZINDO)	eV	0.296	**0.121**
Ionization potential (ZINDO)	eV	0.214	**0.176**
Electron affinity (ZINDO)	eV	0.174	**0.0940**
HOMO (PBE0)	eV	0.155	**0.142**
LUMO (PBE0)	eV	0.129	**0.092**
HOMO (GW)	eV	0.166	**0.142**
LUMO (GW)	eV	0.139	**0.118**
Polarizability (PBE0)	$\mathring{A}^3$	0.173	**0.100**
Polarizability (SCS)	$\mathring{A}^3$	0.149	**0.0578**

4.1 QM7b Dataset

QM7b consists of 7,211 small organic molecules with 14 properties, which is a subset of the GDB-13 database [4]. Each molecule consists of Hydrogen (H), Carbon (C), Oxygen (O), Nitrogen (N), Sulfur (S), and Chlorine (Cl). The three-dimensional coordinates of the most stable conformation and the electronic properties such as HOMO, LUMO, and electron affinity for each molecule are provided calculated by DFT simulation. The discrete graph structure of molecules are missing. Following [13], we train our model per target rather than multi-task learning since the per-target training gives superior performance than joint training.

Table 2 shows mean absolute error (MAE) on QM7b dataset. As shown in the table, our GGRNet consistently outperforms DTNN. The results show that GGRNet is expressive to learn molecular representations as we expected.

4.2 QM8 Dataset

QM8 dataset [22] is also a part of GDB-13 database [4] and contains 21,786 small organic molecules with 12 properties. In QM8, the time-dependent density functional theory (TDDFT) and second-order approximate coupled-cluster (CC2) are applied to calculate the molecular properties. As with QM7b, we train our model per target.

The results are shown in table 3. As shown in the table, our GGRNet achieved the best results on 7/12 cases. However, in some cases such as "f1-CC2" and "f2-CC2", our model gets stuck and suffers from inefficient training. We believe increasing the number of epochs slightly improves the performance, however, more expressive neural architecture is required to learn better molecular representations with a small number of epochs. We hypothesize that tbe gating function of GGR-Net may be the cause of gradient vanishing. One solution is to add batch normalization or other normalization layers to GGRNet.

In "f1" and "f2" dataset, MPNN achieves the best among all other methods. We do not have a clear explanation of the results at present, however, there might be an essential differences between MPNN and GGRNet. One essential difference between MPNN and GGRNet is that MPNN employs expressive readout function by using LSTM, while GGRNet uses simple average function.

---

[1] The MoleculeNet paper incorrectly states this entry as "Excitation energy of maximal optimal absorption - ZINDO"

[2] The MoleculeNet paper incorrectly states this entry as "Highest absorption - ZINDO"

Table 3: QM8 test set performances (MAE)

Property \ Model	GC	DTNN	MPNN	GGRNet
E1-CC2	0.0074	0.0092	0.0084	**0.0057**
E2-CC2	0.0085	0.0092	0.0091	**0.0058**
f1-CC2	0.0175	0.0182	**0.0151**	0.0152
f2-CC2	0.0328	0.0377	**0.0314**	0.0347
E1-PBE0	0.0076	0.0090	0.0083	**0.0053**
E2-PBE0	0.0083	0.0086	0.0086	**0.0054**
f1-PBE0	0.0125	0.0155	**0.0123**	0.0130
f2-PBE0	0.0246	0.0281	**0.0236**	0.0289
E1-CAM	0.0070	0.0086	0.0079	**0.0052**
E2-CAM	0.0076	0.0082	0.0082	**0.0056**
f1-CAM	0.0153	0.0180	**0.0134**	**0.0134**
f2-CAM	0.0285	0.0322	**0.0258**	0.0291

Table 4: QM9 test set performances (MAE)

Property \ Model	Unit	GC	DTNN	MPNN	GGRNet
mu	Debye	0.583	0.244	0.358	**0.172**
alpha	$Bohr^3$	1.37	0.95	0.89	**0.453**
HOMO	Hartree	0.00716	0.00388	0.00541	**0.00372**
LUMO	Hartree	0.00921	0.00513	0.00623	**0.00410**
gap	Hartree	0.0112	0.0066	0.0082	**0.00536**
R2	$Bohr^2$	35.9	17.0	28.5	**1.61**
ZPVE	Hartree	0.00299	0.00172	0.00216	**0.000861**
U0	kcal/mol	3.41	2.43	2.05	31.1
U	kcal/mol	3.41	2.43	2.00	28.0
H	kcal/mol	3.41	2.43	2.02	34.5
G	kcal/mol	3.41	2.43	2.02	30.6
Cv	cal / (mol K)	0.65	0.27	0.42	**0.15**

4.3 QM9 Dataset

QM9 dataset [22] is a widely used comprehensive dataset that provides geometric, energetic, electronic and thermodynamic properties of small organic molecules. It consists of 130k molecules with 12 properties, which are calculated by quantum mechanical simulation method (DFT). Each molecule consists of Hydrogen (H), Carbon (C), Oxygen (O), Nitrogen (N), and Fluorine (F). The number of atoms in each molecule is up to 30. In QM9 dataset, the discrete graph structure of molecules and atom coordinates are provided, while our model does not use the discrete graph structure explicitly. For details about the molecule properties, please refer to [13].

Table 4 shows the experimental results on QM9 dataset. Our GGRNet consistently outperforms other baselines except U0 and U (atomization energy at 0K and 298.15K), H (enthalpy of atomization), and G (free energy of atomization). This results show that our model has potential to learn the representations of small organic molecules in a variety of properties. The energy of the electron in the highest occupied molecular orbital (HOMO) and that of the lowest unoccupied molecular orbital (LUMO) are not improved sufficiently.

Table 5: Ablation study on QM9 dataset.

Model	HOMO	R2	ZPVE
Full model (GGRNet)	0.00372	1.61	0.000861
Full model without $\mathbf{x}_N$	0.00374	2.00	0.000528
Full model without $\{\mathbf{d}_{v,w}\}$	0.0144	154	0.00137
Full model without $\{\mathbf{x}_v\}$	0.00441	1.82	0.000939

4.4 Ablation Study

We performed the ablation study on the subset of QM9 dataset. Table 5 shows the experimental results. The top row is the original GGRNet and the below is the full model without counting feature, distance feature, and atom embedding feature, respectively.

The distance feature is indispensable for the accurate prediction of R2 (an electronic spatial extent). This is also reasonable since this property reflects the spatial distribution of electrons in the molecule. Finally, the atom embedding feature is moderately effective for all properties. Overall, we verified that every additional feature helps to improve the performance for molecular property prediction.

5 Conclusions

In this work, we proposed a GGRNet for accurate and efficient molecular property prediction. In our model, the parameters for updating hidden representations are shared across all layers, the input representations are fed into every layers as skip connections to accelerate the training, and the hidden representation at each atom is updated by using that of every other atoms in the graph, which boosts the performance of the property prediction of molecules. Experiments on the standard benchmarks of molecular property prediction generated by quantum-chemical simulations show that our model achieved the state-of-the-art performance on every datasets. Future work includes applying more expressive functions for update and readout functions for our model.

References

[1] Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, and koray kavukcuoglu. Interaction Networks for Learning about Objects, Relations and Physics. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 4502–4510, 2016.

[2] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993.

[3] Jörg Behler and Michele Parrinello. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Physical Review Letters*, 98(14):146401, 2007.

[4] Lorenz C. Blum and Jean-Louis Reymond. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *Journal of the American Chemical Society*, 131(25):8732–8733, 2009.

[5] Kieron Burke. Perspective on density functional theory. *The Journal of Chemical Physics*, 136(15):150901, 2012.

[6] Raymond E. Carhart, Dennis H. Smith, and R. Venkataraghavan. Atom pairs as molecular features in structure-activity studies: definition and applications. *Journal of Chemical Information and Modeling*, 25(2):64–73, 1985.

[7] R. Qi Charles, Hao Su, Mo Kaichun, and Leonidas J. Guibas. PointNet: Deep Learning on Point Sets for 3d Classification and Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 77–85, Honolulu, HI, 2017. IEEE.

[8] Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017.

[9] Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. Language Modeling with Gated Convolutional Networks. *International Conference on Machine Learning (ICML)*, pages 933–941, 2017. arXiv: 1612.08083.

[10] 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 C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 2224–2232. Curran Associates, Inc., 2015.

[11] Daniel C. Elton, Zois Boukouvalas, Mark S. Butrico, Mark D. Fuge, and Peter W. Chung. Applying machine learning techniques to predict the properties of energetic materials. *Scientific Reports*, 8(1), 2018.

[12] Alex Fout, Jonathon Byrd, Basir Shariat, and Asa Ben-Hur. Protein Interface Prediction using Graph Convolutional Networks. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 6530–6539. Curran Associates, Inc., 2017.

[13] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural Message Passing for Quantum Chemistry. *International Conference on Machine Learning (ICML)*, pages 1263–1272, 2017.

[14] M. Gori, G. Monfardini, and F. Scarselli. A new model for learning in graph domains. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, volume 2, pages 729–734, 2005.

[15] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O. Anatole von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *The Journal of Physical Chemistry Letters*, 6(12):2326–2331, 2015.

[16] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular Graph Convolutions: Moving Beyond Fingerprints. *Journal of Computer-Aided Molecular Design*, 30(8):595–608, 2016.

[17] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated Graph Sequence Neural Networks. *International Conference on Learning Representations (ICLR)*, 2016. arXiv: 1511.05493.

[18] Grégoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New Journal of Physics*, 15(9):095003, 2013.

[19] Kyaw-Zeyar Myint, Lirong Wang, Qin Tong, and Xiang-Qun Xie. Molecular Fingerprint-Based Artificial Neural Networks QSAR for Ligand Biological Activity Predictions. *Molecular Pharmaceutics*, 9(10):2912–2923, 2012.

[20] Xiaojuan Qi, Renjie Liao, Jiaya Jia, Sanja Fidler, and Raquel Urtasun. 3d Graph Neural Networks for RGBD Semantic Segmentation. In *IEEE International Conference on Computer Vision (ICCV)*, pages 5209–5218, Venice, 2017. IEEE.

[21] Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. Semi-supervised User Geolocation via Graph Convolutional Networks. In *Association for Computational Linguistics (ACL)*, pages 2009–2019, 2018.

[22] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1, 2014.

[23] Raghunathan Ramakrishnan, Mia Hartmann, Enrico Tapavicza, and O. Anatole von Lilienfeld. Electronic spectra from TDDFT and machine learning in chemical space. *The Journal of Chemical Physics*, 143(8):084111, 2015.

[24] Bharath Ramsundar, Peter Eastman, Patrick Walters, Vijay Pande, Karl Leswing, and Zhenqin Wu. *Deep Learning for the Life Sciences*. O'Reilly Media, 2019.

[25] David Rogers and Mathew Hahn. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010.

[26] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Physical Review Letters*, 108(5):058301, 2012.

[27] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

[28] Kristof T. Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R. Müller, and Alexandre Tkatchenko. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nature Communications*, 8:13890, 2017.

[29] Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in Neural Information Processing Systems (NIPS)*, pages 991–1001, 2017.

[30] Daniil Sorokin and Iryna Gurevych. Modeling Semantics with Gated Graph Neural Networks for Knowledge Base Question Answering. In *The International Conference on Computational Linguistics (COLING)*, pages 3306–3317, 2018.

[31] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *International Conference on Learning Representations*, 2018.

[32] Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chemical science*, 9(2):513–530, 2018.

[33] Muhan Zhang and Yixin Chen. Link Prediction Based on Graph Neural Networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems (NIPS)*, pages 5165–5175. Curran Associates, Inc., 2018.