Neural Message Passing on High Order Paths

Daniel Flam-Shepherd [1,3*], **Tony Wu** [1*], **Pascal Friederich** [1], **Alan Aspuru-Guzik**[1,2,3]
University of Toronto[1], Canadian Institute for Advanced Research[2], Vector Institute[3]

Abstract

Current graph neural network models do not directly account for local and hidden structures in graphs such as functional groups and molecular geometry in chemical graphs. At each propagation step, most models aggregate only over first order neighbours, ignoring important information contained in subsequent neighbours as well as the relationships between those higher order connections. In this work, we generalize graph neural nets to pass messages and aggregate across paths of higher order. This allows for information to propagate over various levels and substructures of the graph. We experiment with a graph regression task for quantum chemistry calculations and a node classification task in citation networks.

1 Introduction

Graph Neural Networks (GNNs) are a powerful tool for representation learning of relational data across various domains such as chemical compounds [1] or social and biological networks [2]. GNNs are the generalization of convolutional networks to irregular domains. These models learn node embeddings by passing and aggregating node and edge feature information across the graph, typically in a message passing framework [3] using neural networks to pass messages, update node features and readout predictions. The learned node representations can then be used for any downstream procedure such as node or graph classification or regression.

Current GNN models only propagate information across neighbouring edges and – after propagation – use simple pooling of final node embeddings [1, 4]. This means that, in most models, nodes only learn about the larger neighbourhood surrounding them over many propagation steps. This makes it difficult for GNNs to learn higher order graph structure and impossible to learn in a single convolution layer. However, such long range correlations are important for many domains, *e.g.* when learning chemical properties that depend on rings, branches, functional groups or molecular geometry, as well as the relationship between large communities and single nodes in citation networks.

To account for higher order graph properties, we generalize message passing graph neural nets to pass messages and aggregate across simple paths of higher order neighbours. This enables us to use path features in addition to node and edge features, which is very useful in representing 3D information in molecular graphs where many informative features are characterized by the paths between atoms. We experiment with a graph regression task for quantum chemistry calculations and a node classification task in citation networks.

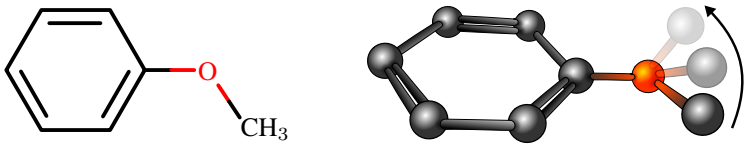

Figure 1: 2D graph and examples of corresponding 3D structures of a molecule.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

2 Background and Related Work

We work with undirected graphs G with n nodes denoted $G = (\mathbf{A}, \mathbf{E}, \mathbf{X})$, where the node adjacency matrix is $\mathbf{A} \in \{0, 1\}^n$ such that $\mathbf{A}_{ij} = 1$ implies nodes i and j are connected. The edge feature tensor $\mathbf{E} \in \mathbb{R}^{n \times n \times e}$ specifies edge features, with $\mathbf{E}_{ijk} = 1$ denoting that nodes i and j have edge type k. The node feature matrix $\mathbf{X} \in \mathbb{R}^{n \times f}$ is a matrix stacking all node feature vectors $\boldsymbol{x}_v \in \mathbb{R}^f$.

Graph Neural Networks and Neural message passing. MPNNs operate on undirected graphs G with node features $\mathbf{x}_v \in \mathbb{R}^f$ and edge features $\mathbf{e}_{vw} \in \mathbb{R}^e$ The forward pass has two phases, a message passing phase and a readout phase. The message passing phase runs for T propagation steps and is defined in terms of message functions M_t and node update functions U_t. During the message passing phase, hidden states $\boldsymbol{h}_v^t$ at each node in the graph are updated based on messages $\boldsymbol{m}_v^{t+1}$ according to

$$\boldsymbol{m}_v^{t+1} = \sum_{w \in \mathcal{N}_v} M_t(\boldsymbol{h}_v^t, \boldsymbol{h}_w^t, \boldsymbol{e}_{vw}), \quad \boldsymbol{h}_v^{t+1} = U_t(\boldsymbol{h}_v^t, \boldsymbol{m}_v^{t+1}), \quad \mathbf{y} = R(\{\boldsymbol{h}_v^T\}_{v \in G})$$

The message node v receives aggregates over its neighbours $\mathcal{N}_v$, in this case, by simple summation. We then readout predictions $\mathbf{y}$ based on final node embeddings.

Higher Order GNNs. Recent work has generalized graph convolution networks (GCNs) [5] to higher order structure by repeatedly mixing feature representations of neighbors at various distances [6], or casting GCNs into a general framework inspired by the path integral formulation of quantum mechanics [7]. Both of these works are based on powers of the adjacency matrix and do not account directly for the relationship between higher order neighbours. Another work [8] proposes k-dimensional GNNs in order to take higher order graph structures at multiple scales into account. Another model based on the transformer architecture [9] accounts for long range dependencies in molecular graphs by augmenting edge feature tensor to include some (shortest) path features like bond type, conjugacy, inter-atomic distance and ring membership.

3 Motivation and Approach

GNNs and higher order GNNs do not incorporate the relationship between higher order neighbours, which would allow for features that are dependent on that relationship, namely 'path features'. An important application is quantum chemistry, where molecular properties are influenced by the geometry of the molecules. The 3D configuration of a molecule can be fully specified by 1) bond lengths – the distance between two bonded atoms, 2) bond angles – the angle formed between three neighbouring atoms, and 3) dihedral angles between four consecutive atoms.

Current GNN approaches to quantum chemistry incorporate neighbouring geometry by using bond distances as edge features [3], but do not directly account for the relative orientation of neighbouring atoms and bonds – a framework that could do so would be advantageous. In general, at every propagation step, GNNs should learn representations that take into account the extended neighbourhood of each node and the features encoded in the relationships between nodes of that neighbourhood.

Our approach. We generalize MPNNs to propagate higher order neighbour information during every propagation step by augmenting them to pass messages along simple paths of higher order neighbours. This allows us to specify path features, which for molecular graphs includes bond angles and dihedral angles – thus encoding the full molecular geometry. We describe the general framework in Section 4.

Model Choice. In this work we focus on two models for two different domains: 1) For node classification in citation networks we use GCNs [5], which is a MPNN with $\boldsymbol{m}_v^{t+1} = \sum_{w \in \mathcal{N}_v} \hat{\mathbf{A}}_{vw} \boldsymbol{h}_v^t$, $U_t = \sigma(\boldsymbol{m}_v^t)$, where σ is a dense layer with sigmoid activation function. 2) For quantum chemistry we use the following MPNN:

$$\boldsymbol{m}_v = \underset{w \in \mathcal{N}_v}{\texttt{Attention}}([\boldsymbol{h}_v^t, \boldsymbol{h}_w, \boldsymbol{e}_{vw}]), \quad U_t = \sigma([\boldsymbol{h}_v^t, \boldsymbol{m}_v^t,]) \quad R = \underset{v \in G}{\texttt{Set2Set}}(\{\boldsymbol{h}_v^T, \boldsymbol{x}_v\})$$

and a variant of MPNN [3] that uses graph attention [10] as an aggregation method and the message function from the model in interaction networks [11], which is a simple concatenation of node and edge features. The node update function concatenates incoming messages with the current node state and feeds it through a dense layer. After propagation through message passing layers, we use the set2set model [12] as the readout function to combine the node hidden features into a fix-sized hidden vector.

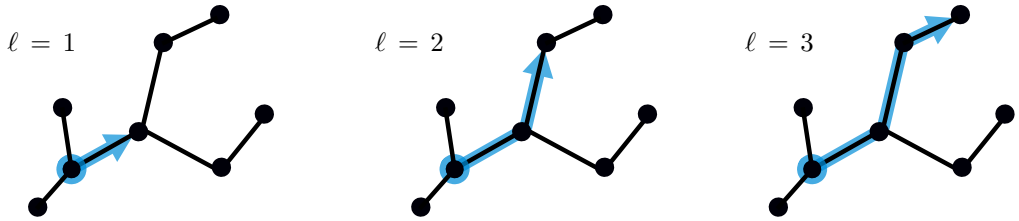

Figure 2: Examples of single messages passed over simple paths of varying length ℓ.

4 Neural Message passing on Paths

We generalize the message passing framework by propagating each node's higher order neighbour information over simple paths instead of aggregating messages from only nearest neighbours. The message passing phase is augmented such that hidden states $\boldsymbol{h}_v^t$ at each node in the graph are updated based on messages over all simple paths up to length ℓ from its neighbours:

$$\boldsymbol{m}_v^{t+1} = \sum_{\boldsymbol{p} \in \mathcal{P}_\ell{}^v} M_t(\boldsymbol{h}_v^t, \boldsymbol{p}) = \sum_{v_1 \in \mathcal{N}_v} \sum_{\substack{v_2 \in \mathcal{N}_{v_1} \\ v_2 \neq v}} \cdots \sum_{\substack{v_\ell \in \mathcal{N}_{v_{\ell-1}} \\ v_\ell \neq v_{\ell-2}, \ldots, v}} M_t(\boldsymbol{h}_v^t, \boldsymbol{p}_{v_1 \to v_\ell}) \tag{1}$$

where we define $\boldsymbol{p}$ to be a path in $\mathcal{P}_\ell{}^v$, which is the set of all simple paths starting from node v with length ℓ and $\boldsymbol{p}_{v_1 \to v_\ell}$ to be path features along path $\boldsymbol{p}$ from node v_1 to node v_ℓ. We only sum over simple paths, excluding loops and multiple inclusions of the same node.

Path sampling and aggregation. For graphs with large number of nodes and edges, passing messages along paths becomes very expensive; for these kinds of graphs, sampling over paths of higher order neighbours is necessary as in GraphSage [2]. Furthermore, in addition to paths, one can aggregate over path features.

Example: Citation Networks. For a citation network the path features are just the node features and edge features connecting v to nodes that are ℓ nodes away, *i.e.*

$$\boldsymbol{p}_{v_1 \to v_\ell} = \{\boldsymbol{h}_{v_1}^t, \boldsymbol{e}_{vv_1}, \ldots, \boldsymbol{h}_{v_\ell}^t, \boldsymbol{e}_{v_\ell v_{\ell-1}}\}$$

Example: Chemical Graphs. In case of molecules, the utility of passing messages over paths becomes very apparent as we explicitly have additional path features depending on path lengths. Considering paths of length 3, where we have a message function that sums over 3 neighbouring atoms. Along with their node and edge features, the path features include bond angles $\boldsymbol{\alpha}_{vwy}$ & $\boldsymbol{\alpha}_{wyx}$ and the dihedral angle $\boldsymbol{\varphi}_{vwyx}$ so that information about the entore local molecular structure is passed over paths of higher order neighbours:

$$\boldsymbol{m}_v^{t+1} = \sum_{w \in \mathcal{N}_v} \sum_{\substack{y \in \mathcal{N}_w \\ y \neq v}} \sum_{\substack{x \in \mathcal{N}_y \\ x \neq w, v}} M_t(\boldsymbol{h}_v^t, \boldsymbol{p}_{v \to x})$$

$$\boldsymbol{p}_{v \to x} = \begin{bmatrix} \boldsymbol{h}_w^t & \boldsymbol{h}_y^t & \boldsymbol{h}_x^t \\ \boldsymbol{e}_{vw} & \boldsymbol{e}_{wy} & \boldsymbol{e}_{yx} \\ \boldsymbol{\alpha}_{vwy} & \boldsymbol{\alpha}_{wyx} & \boldsymbol{\varphi}_{vwyx} \end{bmatrix}$$

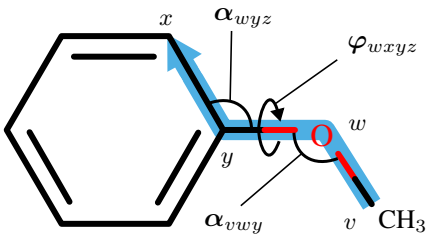

Figure 3: Message function and path features for paths with length 3 in a molecule.

5 Experiments

5.1 Application : Citation Network

For semi-supervised node classification, we benchmark on the CORA dataset which contain sparse bag-of-words feature vectors for each document and a list of citation links between documents which we used as undirected edges to construct the adjacency matrix. Each document has a class label. The network has 2,708 nodes and 5,429 edges with 7 classes and 1,433 features. We use the experimental setup of [5] and benchmark against two other higher order GCN variants: Mixhop and PAN: Path integral graph convolution – both use powers of the adjacency to aggregate GCN layers of higher order neighbours. Our base message passing NN is also a GCN as described in section 3. We sum over paths of length 3 while only sampling a single second order and third order neighbour. Our path features are just node features. Our model achieve similar accuracy to our benchmarks.

Table 1: results on CORA

Model	Test accuracy
GCN [5]	81.5
MixHop [6]	81.9
PAN [7]	82.0
Path GCN	**82.4**

5.2 Application : Quantum Chemistry

Dataset. We benchmark our framework on the QM8 dataset [13, 14], which consists of 21786 organic molecules that have 8 or less heavy atoms (C, O, N and F) and 16 electronic spectra values calculated using density functional theory. The molecules are processed into graphs using the atoms as nodes and bonds as edges. The node and edge features are one-hot encoded atom types (C, O, N, F and H) and bond types (single, double, triple and aromatic bonds).

Training. We split the QM8 dataset into training, validation and test sets with 17786, 2000 and 2000 molecules, respectively. The models are trained using mean squared error (MSE) loss. Model evaluation is done using mean absolute error (MAE) of the molecular properties in the QM8 dataset.

Model architectures. We used the model described in Section 3 cast into the message passing over paths framework, summing over paths of length 3 like example 2) in Section 4. We also compared to a baseline MPNN model that incorporates aggregated angular and dihedral information.

Table 2: Training results on QM8

Model	Test MAE ($\times 10^{-3}$)
GCN-FP [15, 1]	14.80
GGNN [15, 4]	12.67
GCN [15, 5]	11.41
MPNN [15, 3]	11.08
GraphSAGE [15, 2]	12.95
GAT [15, 10]	11.02
LanczosNet [15]	9.58
MPNN (our model)	9.54
Path MPNN	**8.70**

Results. In Table 2, the MAE of our models are compared to the LanczosNet [15] results, which were also benchmarked on the QM8 dataset. Path MPNN achieved the best MAE among all models. To highlight the importance of path features, we also compared performance of training an MPNN model without path features, which also performs better than the LanczosNet, but worse than the Path MPNN.

6 Conclusion

In this work, we introduce a general GNN framework based on message passing over simple paths of higher order neighbours motivated by quantum chemistry. This allows us to use path features in addition to node and edge features, which is very useful in molecular graphs, as many informative features are characterized by the paths between atoms. We benchmarked our framework on a graph regression task for quantum chemistry calculations and a node classification task in citation networks.

References

[1] David Duvenaud, Dougal Maclaurin, Jorge Aguilera-Iparraguirre, Rafael Gómez-Bombarelli, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Neural Information Processing Systems*, 2015.

[2] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

[3] 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-Volume 70*, pages 1263–1272. JMLR. org, 2017.

[4] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493*, 2015.

[5] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[6] Sami Abu-El-Haija, Bryan Perozzi, Amol Kapoor, Hrayr Harutyunyan, Nazanin Alipourfard, Kristina Lerman, Greg Ver Steeg, and Aram Galstyan. Mixhop: Higher-order graph convolution architectures via sparsified neighborhood mixing. *arXiv preprint arXiv:1905.00067*, 2019.

[7] Zheng Ma, Ming Li, and Yuguang Wang. Pan: Path integral based convolution for deep graph neural networks. *arXiv preprint arXiv:1904.10996*, 2019.

[8] Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4602–4609, 2019.

[9] Benson Chen, Regina Barzilay, and Tommi Jaakkola. Path-augmented graph transformer network. *arXiv preprint arXiv:1905.12712*, 2019.

[10] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[11] Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, et al. Interaction networks for learning about objects, relations and physics. In *Advances in neural information processing systems*, pages 4502–4510, 2016.

[12] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets. *arXiv preprint arXiv:1511.06391*, 2015.

[13] Lars Ruddigkeit, Ruud Van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of chemical information and modeling*, 52(11):2864–2875, 2012.

[14] 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.

[15] Renjie Liao, Zhizhen Zhao, Raquel Urtasun, and Richard S Zemel. Lanczosnet: Multi-scale deep graph convolutional networks. *arXiv preprint arXiv:1901.01484*, 2019.