A Theoretical Formulation of Many-body Message Passing Neural Networks

Jiatong Han [1]

Abstract

We present many-body Message Passing Neural Network (MPNN) framework that models higher-order node interactions (≥ 2 nodes). We model higher-order terms as tree-shaped motifs, comprising a central node with its neighborhood, and apply localized spectral filters on motif Laplacian, weighted by global edge Ricci curvatures. We prove our formulation is invariant to neighbor node permutation, derive its sensitivity bound, and bound the range of learned graph potential. We run regression on graph energies to demonstrate that it scales well with deeper and wider network topology, and run classification on synthetic graph datasets with heterophily and show its consistently high Dirichlet energy growth.

We open-source our code at https://github.com/JThh/Many-Body-MPNN.

1. Introduction

We study a generic graph setting where no information on distances between nodes or rotations of edges is available. We construct many-body message to increase the receptive field of a single message-passing step, than the two-body case, as explicitly encoding many-body interaction reduces the need of stacking layers for message-passing (Batatia et al., 2022).

Prior work such as ChebNet (Defferrard et al., 2016) approximates the multi-hop messages from neighbors through polynomial expansions to avoid directly computing the powers of adjacency matrices. However, the spectral filters of ChebNet are making signal transformations to graphs globally, despite its efforts to localize spectral filtering through Chebyshev expansion via finite support size k. We apply a simple yet effective localization by explicitly defining a series of motifs (Monti et al., 2018) where smaller-scale spectral filters are applied.

We modify motif Laplacian transform to distinguish the contribution from each node in the motif. This involves mapping the feature vectors to the Fourier basis of the concerned motif and scaling by the edge's curvature, before mapping them back to graph domain. We enumerate the neighboring sets of varied sizes and symmetrically aggregating the filtered outcomes to ensure permutation invariance.

Our contributions are to: 1) formulate many-body message-passing, prove its invariance to neighbor node permutation, and make explicit the contribution of every body-order interaction (Equation (6)); 2) derive the new sensitivity and energy bound of the many-body interaction paradigm (Equation (7), Theorem 5.2); 3) show many-body MPNN scale well with wider and deeper network topology and achieve significant energy growth on heterophilic graphs (Section 6.2 and 6.3).

2. Preliminary

We study general graph structure G that have node set N and edge set E of finite sizes. We use $e_{i,j}$ to denote edge between node i and j. We use D for degree matrix, A for adjacency matrix, $L = (D - A)$ for unnormalized graph Laplacian, $\mathcal{L} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ as symmetric normalized Laplacian. We define correlation order ν to be the order to which we model the body interactions explicitly in our framework, including the central node.

For different correlation orders, we use motifs (undirected graph sub-structures) to represent the local graph formed by the central node and its neighbors in consideration. We define k-motifs to specifically refer to undirected tree graphs with one central node and $(k-1)$ branches to localize many-body interactions of different orders. The adjacency matrix (and hence Laplacian L) of motifs can be assigned weights such as edge curvatures for localization. We rely on Balanced Forman curvature, a combinatorial definition of Ricci curvature as established in Topping et al. (2022, Def. 1), and we hereafter refer to it as Ricci curvature and the function $\text{Ricci}(e_{i,j})$ for curvature of edge $e_{i,j}$.

[1]Department of Computer Science, University of Oxford, United Kingdom. Correspondence to: Jiatong Han <jiatong.han@cs.ox.ac.uk>.

Proceedings of the Geometry-grounded Representation Learning and Generative Modeling at 41st International Conference on Machine Learning, Vienna, Austria. PMLR Vol Number, 2024.
Copyright 2024 by the author(s).

Potential function. A graph's potential $\mathcal{E}$ can be measured with Dirichlet energy. Dirichlet energy is defined as $Tr(\boldsymbol{X}^T\mathcal{L}\boldsymbol{X}) = \sum_{(i,\ j)\in\mathsf{E}} \left\|\boldsymbol{X}_i/\sqrt{d_i} - \boldsymbol{X}_j/\sqrt{d_j}\right\|_2^2$, where $\mathcal{L}$ is the normalized Laplacian, $\boldsymbol{X}$ is the set of node features, and d's are node degrees.

3. Formulation of Many-body MPNN

The generalized message construction involving all correlation orders ν[1] as defined by Batatia et al. (2022) is :

$$m_i^{(t)} = \sum_j \nu_1(\sigma_i^{(t)}, \sigma_j^{(t)}) + \sum_{j_1,j_2} \nu_2(\sigma_i^{(t)}, \sigma_{j_1}^{(t)}, \sigma_{j_2}^{(t)}) + \cdots, \tag{1}$$

where j's are node i's neighbors and $\sigma_j^{(t)}$ is the learned representation of node j for computing ν-th correlation strength at layer t.

For modelling interactions of different body orders, we follow the format of Equation (1) and avoid explicit message passing via spectral filters. We start by computing the two-body interaction messages using graph convolution. This involves apply graph Fourier transform to neighboring nodes, weighting their contributions via learnable parameters, and localizing the signal transformation through second-order Chebyshev expansion.

The two-body interaction component of our model captures the pairwise relationships between nodes. By incorporating node features from previous iterations, we compute the interaction messages as follows:

$$\boldsymbol{X}^{(t)} = \boldsymbol{U}^\top g_{\theta_2}(\boldsymbol{\Lambda})\boldsymbol{U}\boldsymbol{H}^{(t-1)}, \tag{2}$$

where $\boldsymbol{X}^{(t)}$ represents the two-body interaction matrix at iteration t, $\boldsymbol{U}$ is the matrix of eigenvectors of G's normalized Laplacian. The function $g_{\theta_2}(\boldsymbol{\Lambda})$ denotes a second-order Chebyshev polynomial expansion applied to the diagonal matrix $\boldsymbol{\Lambda}$, which holds the eigenvalues of the Laplacian matrix of the graph G.

Higher-order interaction message (Equation (3)) captures the complex interactions involving more than two nodes, utilizing motif-based structures for higher-order relations. We draw on the intuition from Windowed Fourier Transform (WFT) to apply graph signal filtering to subgraphs, specifically motifs sized according to the correlation order. These motifs are standardized undirected subgraphs centered around a node with neighborhoods of size (- 1). The output message for each node maintains permutation invariance by explicitly enumerating (by η_ν) all neighboring sets of size $k \leq \nu$.

To better locate each motif in global graph G, we modify the motif Laplacian to account for Ricci curvature. Edges with more negative Ricci curvature are assigned higher positive weights in the local Laplacian of G's motif, reflecting more critical connections in the graph structure:

$$L_{\text{Ricci-J}_{(i,j)}} = \begin{cases} \sum_{j\neq i} \text{Ricci}(e_{i,j}) & \text{if } i = j \\ -\text{Ricci}(e_{i,j}) & \text{if } i \neq j \\ 0 & \text{otherwise,} \end{cases}$$

where J is the neighbor node set of node i and $j \in J$.

The higher-order message $\boldsymbol{Y}^{(t)}$ is hence formulated as (with $\nu \geq 3$):

$$\boldsymbol{Y}_i^{(t)} = \prod_{k=3}^{\nu} \sum_{\substack{J\subseteq\eta(N(i))\cup\{i\},\\ |J|=k}} \boldsymbol{U}_k^\top g_{\theta_k}(\boldsymbol{\Lambda}_{\text{Ricci-J}})\boldsymbol{U}_k\boldsymbol{H}_J^{(t-1)}, \tag{3}$$

where η is the enumeration of element sets from the neighbors of node i. $U_k^\top$ is the inverse or transpose of eigenvectors of $L_{\text{Ricci-J}}$. And $g_{\theta_k}(\boldsymbol{\Lambda}_{\text{Ricci-J}})$ term is defined as:

$$g_{\theta_k}(\boldsymbol{\Lambda}_{\text{Ricci-J}}) = \sum_{k'=1}^{k} \theta_{k,k'} T_{k'}(\widetilde{\boldsymbol{\Lambda}_{\text{Ricci-J}}}), \tag{4}$$

where $\widetilde{\boldsymbol{\Lambda}_{\text{Ricci-J}}} = \frac{2\boldsymbol{\Lambda}_{\text{Ricci-J}}}{\lambda_{\text{max-J}}} - I_{|J|}$ and $\boldsymbol{\Lambda}_{\text{Ricci-J}}$ is the diagonal matrix of eigenvalues of $|J|$-motif's Laplacian $L_{\text{Ricci-J}}$ whose weights are obtained from the edge set $\mathsf{E}_{i,J} := \{e_{i,j} \mid j \in J\}$ with respect to G, with $\lambda_{\text{max-J}}$ representing the largest eigenvalue of the Laplacian $L_{\text{Ricci-J}}$.

And the feature update equation, which updates the node features by combining the self-features and aggregated neighbor messages, establishes a residual connection:

$$\begin{aligned} \boldsymbol{h}_i^{(t)} &= \boldsymbol{h}_i^{(t-1)} + \boldsymbol{W}^{(t)} m_i^{(t)} \\ &= \boldsymbol{h}_i^{(t-1)} + \boldsymbol{W}_x^{(t)}\boldsymbol{X}_i^{(t)} + \boldsymbol{W}_y^{(t)}\boldsymbol{Y}_i^{(t)} \end{aligned} \tag{5}$$

4. Many-Body Mixing Bound: Sensitivity Bound

For node i at layer t, the update function of all body orders can be formulated as:

[1] Here we have overloaded the notation ν to denote many-body interaction function, aside from its initial meaning of correlation orders.

$$\boldsymbol{h}_i^{(t)} = \sum_{k'=0}^{2} \theta_{2,k'}^{(t)} T_{k'}(\widetilde{\mathcal{L}}) \boldsymbol{h}_i^{(t-1)} + \prod_{k=3}^{\nu} \sum_{J \subseteq \eta_k(N(i))} \sum_{k'=1}^{k} \theta_{k,k'}^{(t)} T_{k'}(\widetilde{L_{\text{Ricci-J}}}) \boldsymbol{H}_J^{(t-1)} \tag{6}$$

, where $\boldsymbol{h}_i^{(0)} = \boldsymbol{x}_i$, with the residual term included when $k' = 0$.

The over-squashing effect can be understood with node's representation $\boldsymbol{h}_u^{(r)}$ failing to be affected by some input feature $\boldsymbol{x}_v$ of node v at distance r from node u. We use the Jacobian ($\partial \boldsymbol{h}_u^{(r+1)} / \partial \boldsymbol{x}_v$) to assess over-squashing and derive sensitivity bounds, similar as Topping et al. (2022).

Theorem 4.1. *The sensitivity bound of a many-body MPNN with update function as defined in Equation (6) is given by*

$$O\left(\left|\frac{\partial \boldsymbol{h}_u^{(r+1)}}{\partial \boldsymbol{x}_v}\right|\right) = O\left((\mathcal{A}^r)_{vu} \boldsymbol{x}_v^{\nu^r}\right) \tag{7}$$

where $\nu \geq 2$ is the correlation order, r is the shortest distance between node v and node u.

The powers of x_v in Equation (7) come from the product term $\prod$ from Equation (3), different from an MPNN's sensitivity bound that does not usually involve node initializations (Giovanni et al., 2024), as they are negligible constants.

5. Properties of Many-Body MPNN

5.1. Invariance Property

We reconstruct message construction as $\mathcal{F}_m(\boldsymbol{H}^{(t-1)}, L) = W^{(t)} m^{(t)}$ from Equation (5).

Theorem 5.1. *The message construction function ($\mathcal{F}_m$) is invariant to the permutation of neighbors $N(i)$ of input nodes $i \in \mathsf{N}$, considering different body numbers ν, i.e.*

$$\mathcal{F}_m(\boldsymbol{H}^{(t-1)}, L) = \mathcal{F}_m(\boldsymbol{H}^{(t-1)}, \pi^\top L \pi), \tag{8}$$

assuming π is the permutation matrix.

5.2. Energy Bounds

Theorem 5.2. *Many-body MPNN with finite number of layers (t) learns the graph potential ($\mathcal{E}$) into a bounded range:*

$$\left[0, \quad \lambda_{max} |\mathsf{N}| \left(\prod_{\nu} \nu^t \binom{d_{max}}{\nu - 1}^t \prod_t \boldsymbol{w}_\nu^{(t)} \boldsymbol{h} \right)^2 \right],$$

where d_{max} is the maximum node degree and correlation order $\nu \geq 2$, and $\boldsymbol{w}^{(t)}, \boldsymbol{h}$ are assumed constant upper bounds for weights of layer t, and the initial node features.

We derive from Theorem 5.2 that many-body MPNN has higher energy upper-bound than ChebNets, and higher-order terms produce strictly more energy than lower-order terms, given the same number of layers t. See Appendix A.3.

5.3. Complexity Bounds

Proposition 5.3. *There exists a simplified implementation of many-body MPNN whose runtime is linear in $(|\mathsf{E}| + |\mathsf{N}|)$ and has time complexity $O(2|\mathsf{E}| + |\mathsf{N}|(d_{max})^{\nu-1})$ per layer, where d_{max} is the maximum node degree and correlation order $\nu \geq 2$.*

We prove in Proposition 5.3 that despite the additional complexity of constructing higher-order messages, the many-body implementation can execute at comparable speed as the two-body case in its optimal state. See Section 6.4.

6. Experiments

We demonstrate the proven properties of many-body MPNN through graph regression and node classification tasks, illustrating its capability to scale with more extensive and deeper architectures and to capture complex local node interactions.

6.1. Experimental Settings.

Due to the sheer computational complexity of approximating Ricci curvatures and weighted motif Laplacian, we follow Proposition 5.3 to simplify our implementation. We compare our model performances with two-body MPNNs and other graph convolutional networks such as GCNs (Kipf & Welling, 2017) and ChebNets (Defferrard et al., 2016).

6.2. Regressing Synthetic Random Graph Energies

We experiment with random graphs and synthesized energy functions. We design our energy functions through applying non-linear transformations such as exponential or logarithmic functions on graph average shortest path lengths or average clustering coefficient, to emphasize either distant node mixing or local node clustering.

We experiment with 100 Erdős–Rényi Graph having 500 to 700 nodes with edge probabilities 0.15 to 0.3. We observe in Figure 1 that when emphasizing node distances, stacking more many-body MPNN layers or decreasing hidden dimensions let it perform observably better than others, yet **the opposite** is witnessed when emphasizing local clustering, which we understand as wider neural networks benefit learning more complex local node interactions.

6.3. Classifying Synthetic Heterophilic Graph Nodes

We synthesize graph datasets for node classification task that exhibits highly heterophilic properties to illustrate graph

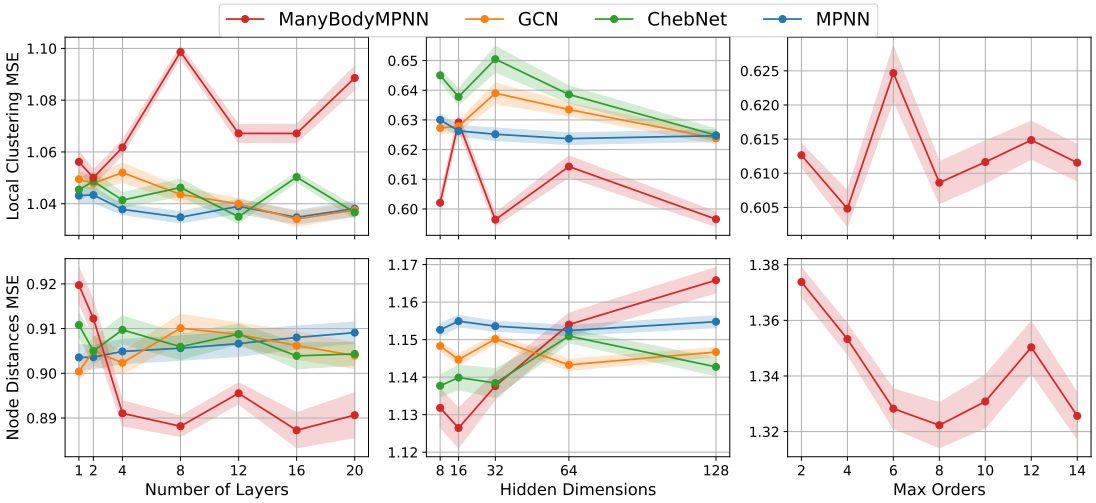

Figure 1. Test MSE Losses on 100 synthetic random graphs. We show the impact of varying number of layers, hidden dimensions, and max correlation orders on the energy regression. The upper row is regressing graph energies emphasizing clustering, with the lower emphasizing node distances.

learning capabilities. We generate a single heterophilic graph with 10,000 nodes with 7 classes, each of which has 1,433 dimensional features, with an average node degree of 10, as a larger replacement of Cora (Chen et al., 2018). We measure the test accuracy and the Dirichlet energy growth over epochs (see Figure 2).

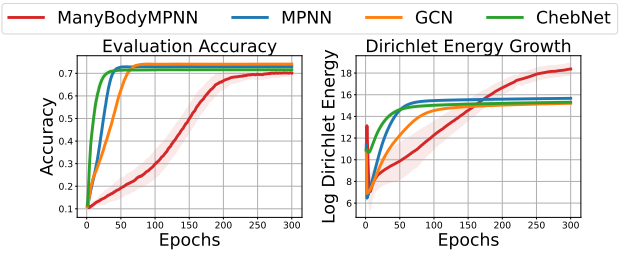

Figure 2. Test accuracy and Dirichlet energy growth of models over 300 epochs on a synthetic heterophilic graph.

Many-body MPNN generates much higher energy than other networks, despite its slower convergence, signalling its capability of learning the representations of contrasting nodes quite differently.

6.4. Efficiency Benchmarks

We benchmark model speeds for different layer counts in Figure 3. We observe that with runtime increasing linearly with layer counts, many-body MPNN runs 3.44 times slower than ChebNet when having 20 layers, mainly attributed to higher-order interactions. We may further improve the running time as proposed in Proposition 5.3 in future efforts.

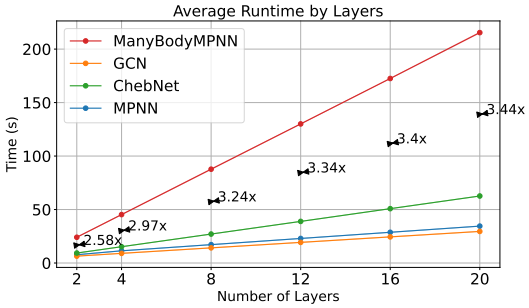

Figure 3. Runtime for different models, averaged over 30 runs. The models have varying numbers of layers, 16 hidden dimensions, and 4 correlation orders. The support for the Chebyshev expansion is 4. The batch size is 4, and the number of epochs is 50. Benchmarked on one NVIDIA RTX 2080 Ti GPU.

7. Discussion

The effectiveness of many-body MPNN for downstream tasks. Our preliminary results show that many-body formulation scales well with increasing network depths when modeling the distant node interactions, and with increasing network widths when modeling complex local node interactions. It captures both aspects which are not easily learned through stacking more MPNN layers due to over-squashing (Giovanni et al., 2024), or introducing wider layers.

Localized spectral filtering and efficiency concerns. ChebNet applies signal processing to the entire graph, while our approach is more localized by applying to motifs while accounting for the edge contributions through its curvatures. We make the contribution of every sized motif ex-

plicit, and capture more subtle signals that only exists locally. One direct benefit of our approach is the higher energy upper-bound (Theorem 4.1, Corollary A.2); on learning heterophilic graph node labels, it generates more energy over learning for contrasting node labels than ChebNet through its many-body interactions, despite its slightly harder convergence with its heavy parameterization, and we leave its improvement for future work.

Towards learnable embeddings. For input graphs with edge features, the Ricci curvatures can actually be replaced with learnable parameters (Batatia et al., 2022). Additionally, depending on what other information is available in the graphs, such as node positional encoding, distances or commute time, etc. (Black et al., 2023), we may replace Ricci curvatures with these metrics to make their impact more explicit, which might open up exciting avenues for many-body MPNN research.

8. Related Work

MACE. The MACE (Message passing neural network for Atom-Centered Potentials) framework (Batatia et al., 2022), symbolized by the correlation order ν, predicts atomistic potentials within molecular structures. It adapts to the permutation invariance inherent in molecules, due to the indistinguishability of atoms of the same element, and ensures energy conservation. MACE can be expressed by:

$$E(\mathbf{r}^N) = \sum_{i=1}^{N} \nu_i(\mathbf{r}_i, \mathbf{r}_i^{'}, \ldots, \mathbf{r}_i^{(\nu)}) \tag{9}$$

where $E(\mathbf{r}^N)$ represents the potential energy of a system with N atoms, $\mathbf{r}_i$ denotes the position of the i-th atom, and $\mathbf{r}_i^{(\nu)}$ denotes its ν-th order interaction.

Graph Convolutional Neural Networks. ChebNet (Defferrard et al., 2016) introduces localized spectral filters within graph neural networks. The spectral filters are based on Chebyshev polynomials, $T_k(x)$, which serve as an efficient approximation to the graph Laplacian's eigendecomposition, facilitating faster convolutions. For a graph signal $\mathbf{x}$ and a filter g_θ, the operation in ChebNet can be formulated as:

$$\mathbf{x} * g_\theta = \sum_{k=0}^{K-1} \theta_k T_k(\widetilde{\mathbf{L}})\mathbf{x} \tag{10}$$

where $*$ denotes the convolution operation, θ is a vector of Chebyshev coefficients, $\widetilde{\mathbf{L}}$ is the scaled Laplacian, and K represents the order of the polynomial approximation.

Understanding Over-Squashing via Graph Curvature (Giovanni et al., 2023). We adapt from the Balanced Forman curvature formulation from Giovanni et al. (2023) to weigh the edge importance in higher-order message construction. We derive the sensitivity bound (Theorem 4.1) based on the bound for two-body MPNN from Giovanni et al. (2023, Lemma 1), and demonstrate the many-body formulation is capable to scale to deeper network topology.

Understanding Graph Convolutions via Energies (Rusch et al., 2023; Topping et al., 2022; Chamberlain et al., 2021). We leverage the concept of Dirichlet energy and measure how well models embed node differently enough in graphs with heterophily. While we do not ablate model layers while doing node classification, we see many-body MPNN is less likely to hit the energy flow bottleneck than other models through its high energy generation.

9. Conclusion

In this paper, we develop a theoretical formulation of many-body MPNN, which models higher-order node interactions beyond traditional two-body interactions. We address ChebNet's limitations in localizing graph transform effects by designing a technique that enumerates tree-shaped motifs for each correlation order and applies Chebyshev-expanded spectral filters. We derive sensitivity and energy bounds for many-body MPNN and evaluate its performance on synthetic graph energy regression and heterophilic node classification tasks. Our results show that many-body MPNN scales effectively with network depth and width, produces high energy, and maintains test accuracy on par with other convolutional networks, while exhibiting reduced susceptibility to either over-squashing or over-smoothing.

Acknowledgement

We acknowledge the contribution of ideas from Francesco Di Giovanni (francesco.di.giovanni@valencelabs.com) and Michael Bronstein (michael.bronstein@cs.ox.ac.uk) during their time at the University of Oxford. The paper was extended from a course (Geometric Deep Learning) assessment paper solely written by the author.

References

Batatia, I., Kovács, D. P., Simm, G. N., Ortner, C., and Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. volume 35, 2022.

Black, M., Wan, Z., Nayyeri, A., and Wang, Y. Understanding oversquashing in gnns through the lens of effective resistance. volume 202, 2023.

Chamberlain, B. P., Rowbottom, J., Gorinova, M., Webb, S., Rossi, E., and Bronstein, M. M. Grand: Graph neural diffusion. volume 139, 2021.

Chen, J., Ma, T., and Xiao, C. Fastgcn: Fast learning with graph convolu-tional networks via importance sampling. 2018.

Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. 2016.

Giovanni, F. D., Giusti, L., Barbero, F., Luise, G., Liò, P., and Bronstein, M. On over-squashing in message passing neural networks: The impact of width, depth, and topology. volume 202, 2023.

Giovanni, F. D., Rusch, T. K., Bronstein, M. M., Deac, A., Lackenby, M., Mishra, S., and Veličković, P. How does over-squashing affect the power of gnns?, 2024.

Kingma, D. P. and Ba, J. L. Adam: A method for stochastic optimization. 2015.

Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. 2017.

Monti, F., Otness, K., and Bronstein, M. M. Motifnet: A motif-based graph convolutional network for directed graphs. 2018. doi: 10.1109/DSW.2018.8439897.

Rusch, T. K., Bronstein, M. M., and Mishra, S. A survey on oversmoothing in graph neural networks, 2023.

Topping, J., Giovanni, F. D., Chamberlain, B. P., Dong, X., and Bronstein, M. M. Understanding over-squashing and bottlenecks on graphs via curvature. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022.

A. Proofs

A.1. Discussion of Theorem 4.1

The hidden feature $\boldsymbol{h}_i^{(t)}$, computed by a many-body MPNN with t layers as in Equation (6) is a differentiable function of the input node features $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$.

We simplify Equation (6) into more plain expressions in Equation (11). We denote parameter tensors multiplied with constants (including edge curvatures) as $\boldsymbol{c}_{b,d}^{(a)}$'s for brevity.

$$\boldsymbol{h}_i^{(t)} = \boldsymbol{c}_0^{(t)} \boldsymbol{h}_i^{(t-1)} + \prod_{k=2}^{\nu} \left(\boldsymbol{c}_{k,1}^{(t)} \boldsymbol{h}_{o(1)}^{(t-1)} + \boldsymbol{c}_{k,2}^{(t)} \boldsymbol{h}_{o(2)}^{(t-1)} + \cdots + \boldsymbol{c}_{k,|N(i)|}^{(t)} \boldsymbol{h}_{o(|N(i)|)}^{(t-1)} \right)$$

, where $o(j)$ is used to denote the identity of j-th neighbor of node i.

With Equation (11), we may relate the shortest path $\bar{p}$ to their sensitivity bounds. Assume $\mathcal{P}(v, u)$ is the set of paths connecting nodes v and u, and $\ell(p)$ as length of path $p \in \mathcal{P}$, with $\ell(\bar{p}) = r$. Let p_i be the i-th node along path p and $\boldsymbol{S}_{p_i}^{(r)} = (\sum_{j \in N(p_i)} \boldsymbol{h}_j^{(r)})$ be the direct sum of neighborhood of node p_i at layer r. We may derive $\boldsymbol{h}_u^{(r+1)}$ by induction. The first term $\boldsymbol{h}_{p_1}^{(1)}$ along the path that contains x_v terms can be derived as:

$$\boldsymbol{h}_{p_1}^{(1)} = \sum_{k=1}^{\nu-1} \boldsymbol{c}_{p_1,k} \cdot (\boldsymbol{x}_v)^k \cdot (\boldsymbol{S}_{p_1}^{(0)})^{(\nu-k)} \tag{11}$$

And $\boldsymbol{h}_u^{(r+1)}$ can be inductively derived as:

$$\boldsymbol{h}_u^{(r+1)} = \sum_{p \in \mathcal{P}(v,u)} \sum_{k=1}^{\nu-1} \boldsymbol{c}_{p_r,k} \cdot (\boldsymbol{h}_{p_r}^{(r)})^k \cdot (\boldsymbol{S}_{p_r}^{(r)})^{(\nu-k)} \tag{12}$$

By induction, and the fact that $\bar{p} \in \mathcal{P}(v, u)$ can be obtained from powers of symmetrically normalized adjacency matrix $\mathcal{A}$ and the gradients from other paths would not have reached u after r propagations, we may finally derive the bound of $|\partial \boldsymbol{h}_u^{(r+1)} / \partial \boldsymbol{x}_v|$ as in Theorem 4.1.

A.2. Proof of Theorem 5.1

Proof. We assume that node features are initialized uniformly across graphs. We proceed by inducting on $\nu \geq 2$ (assuming that there are at least 2 bodies involved in message construction).

When $\nu = 2$, $m^{(t)} = X^{(t)}$. From Equation (2), we may have the permutation matrix moved inside the Laplacian's eigen-decomposition:

$$\mathcal{F}_m(\boldsymbol{H}^{(t-1)}, \pi^\top L \pi) = W^{(t)} \boldsymbol{U}^\top \left(\pi^\top g_{\theta_2}(\boldsymbol{\Lambda}) \pi \right) \boldsymbol{U} \boldsymbol{H}^{(t-1)} \tag{13}$$

Since Λ is a diagonal matrix and $g_{\theta_2}(\boldsymbol{\Lambda})$ is linearly parameterizing the Chebyshev expansion of Λ, we may confirm that $\pi^\top g_{\theta_2}(\boldsymbol{\Lambda}) \pi$ is an actual permutation of original outcomes. Due to the assumption that node features are initialized the same, final weighted outcome by $W^{(t)}$ will also be the same, and **by induction**, $\mathcal{F}_m(\boldsymbol{H}^{(t-1)}, L) = \mathcal{F}_m(\boldsymbol{H}^{(t-1)}, \pi^\top L \pi)$, which preserves permutation invariance.

For $\nu \geq 3$, the message construction includes the higher order term $Y^{(t)}$ from Equation (3). Due to the fact that permutation of node identities will not alter the graph topology, for each enumerated neighbor set $J \in \eta(N(i))$, the edge set $\mathsf{E}_{i,J}$'s Ricci curvatures will not change. Hence, the weighted motif's Laplacian $\mathcal{L}_{\text{Ricci-J}}$ and its eigenvalues will be unaltered. With a fair η, each neighbor node should have equal probabilities to hold any position (in the local motif's topology) of any motif sizes. Hence permutation of node identities will not affect the message construction. □

A.3. Proof of Theorem 5.2

Proof. We first prove the bound for a single-layer case, and then generalize the result to any number of layers. We assume that the node features at layer 0 for node $i \in \mathsf{N}$ are initialized with $X_i = |\boldsymbol{H}_i^{(0)}| \leq \boldsymbol{h}$. We assume linear weights are initialized with $|\theta_k| \leq \boldsymbol{w}_k$.

We analyze the two-body $\boldsymbol{X}$ and many-body interaction term $\boldsymbol{Y}$ for node features $\boldsymbol{H}^{(1)}$ to approximate the bounds of $\mathcal{E}$ of graph after applying Equation (5)'s updates.

For two-body interaction term $\boldsymbol{X}$, we leverage the fact that graph Laplacian $\mathcal{L}$ is symmetric and positive semi-definite, and hence its eigenvalues are bounded by $[0, \lambda_{\text{max}}]$. After shifting its eigenvalues to be $\tilde{\boldsymbol{\Lambda}}$ as in Equation (4), and since Chebyshev polynomials applied on $[-1, 1]$ are bounded by $[-1, 1]$, it is straightforward that

$$\boldsymbol{X}_i^{(1)} \in [-d_{\text{max}}\boldsymbol{w}_2\boldsymbol{h}, d_{\text{max}}\boldsymbol{w}_2\boldsymbol{h}], \qquad \text{for } i \in \mathsf{N} \tag{14}$$

For many-body interaction term $\boldsymbol{Y}$, there are finite number of neighboring sets for each node i to form unique motifs, and with similar analysis,

$$|\boldsymbol{Y}_i^{(1)}| \leq \prod_3^{\nu} \nu \binom{d_{\text{max}}}{\nu - 1} \boldsymbol{w}_\nu \boldsymbol{h}, \qquad \text{for } \nu \geq 3, i \in \mathsf{N}. \tag{15}$$

Hence, $H_i^{(1)} \in [-\prod_\nu \nu \binom{d_{\text{max}}}{\nu-1} \boldsymbol{w}_\nu \boldsymbol{h}, \prod_\nu \nu \binom{d_{\text{max}}}{\nu-1} \boldsymbol{w}_\nu \boldsymbol{h}]$, after we combine the two interaction terms.

Then the graph potential is computed and bounded as

$$\mathcal{E} := H^{(1)T} \mathcal{L} H^{(1)} \in \left[0, \lambda_{\text{max}} |\mathsf{N}| \left(\prod_\nu \nu \binom{d_{\text{max}}}{\nu - 1} \boldsymbol{w}_\nu \boldsymbol{h} \right)^2 \right], \tag{16}$$

where the maximum is obtained when $H^{(1)}$ aligns with the eigenvector corresponding to λ_{max} and is at its maximal magnitude, and the minimum is zero, when graph nodes have the same embeddings.

We generalize the result to $t \geq 2$ layers. It can be shown by simple induction that $|H^{(t)}| \leq \prod_\nu \nu^t \binom{d_{\text{max}}}{\nu-1}^t \prod_t \boldsymbol{w}_\nu^{(t)} \boldsymbol{h}$.

And the energy bound is hence $\left[0, \lambda_{\text{max}} |\mathsf{N}| \left(\prod_\nu \nu^t \binom{d_{\text{max}}}{\nu-1}^t \prod_t \boldsymbol{w}_\nu^{(t)} \boldsymbol{h} \right)^2 \right]$. □

Lemma A.1. *Higher-order interaction term generates more energy than lower-order terms.*

Proof. With the same number of layers t, node and linear weight bounds ($\boldsymbol{h}$ and $\boldsymbol{w}$), the remaining higher order term $\nu(d_{\text{max}})^{\nu-1}$ grows monotonically with ν. □

Corollary A.2. *ChebNet has strictly lower energy upper-bound than many-body MPNN, with a single layer.*

This is a corollary to Theorem 5.2 since many-body formulation is identical to ChebNet when $\nu = 2$, and the maximum energy contributed by many-body interaction term Y is strictly positive (and higher by A.1).

A.4. Proof of Proposition 5.3

We will prove the time complexity upper-bound to be $O(|\mathsf{E}|d_{\text{max}}^2 + 2|\mathsf{E}| + |\mathsf{N}|(3d_{\text{max}})^{\nu-1}\nu^3)$, and discuss the result in paper on a separate note, following this proof.

Proof. There are three major parts of computation within a many-body MPNN update function: Balanced Forman curvatures for G, two-body interaction term, and many-body interaction term.

Computing the Balanced Forman curvature on G takes $O(|\mathsf{E}|d_{\text{max}}^2)$, according to the formulation in Topping et al. (2022). The curvature value for each edge in E is pre-computed for constructing each motif's weighted Laplacian matrix.

For two-body interaction term, it has the same time complexity as ChebNet (Defferrard et al., 2016), which is $O(|\mathsf{E}| \cdot K)$ where $K = 2$ is the expansion order of Chebyshev polynomials from Equation (2).

For many-body interaction term, we make a simplification that Balanced Forman curvatures (ranging from -2 to positive infinity) are rounded into $\{-1, 0, 1\}$ by the sign of continuous curvature values. The number of uniquely weighted motif's Laplacian of correlation order ν is:

$$\#\mathcal{L}_{\text{motif-J}} \coloneqq \sum_{i \in \mathsf{N}} \binom{d_i}{\nu - 1} 3^{\nu-1} \tag{17}$$

Since eigen-decomposition of Laplacians of shape (ν, ν) takes $O(\nu^3)$ time, Equation (3) takes $O(|\mathsf{N}|(3d_{\text{max}})^{\nu-1}\nu^3)$. And the overall time complexity of simplified many-body MPNN per layer is $O(|\mathsf{E}|d_{\text{max}}^2 + 2|\mathsf{E}| + |\mathsf{N}|(3d_{\text{max}})^{\nu-1}\nu^3)$.

□

On a side note, we may further simplify many-body MPNN formulation through pre-computing the eigen-decomposition of motif's unweighted Laplacian, and instead of assigning edge curvatures as Laplacian weights, we learn the motif edge contribution through learnable parameters specific to each correlation order, from the Chebyshev expansion process (Equation (4)). Since there is no need to compute the curvature values, the time complexity of this further simplified many-body MPNN is $O(2|\mathsf{E}| + |\mathsf{N}|(d_{\text{max}})^{\nu-1})$.

B. Experiments

B.1. Details on Experiment Configurations

For heterophilic graph node classification, We use 4 convolutional layers; we set correlation order ν to be 5 for many-body MPNN and Chebyshev filter size to be 3 for ChebNet. For synthetic spine graph mixing power experiments (Appendix B.2), we train many-body MPNN with layer number in $[4, 8, 13, 25]$, widths in $[8, 32, 128, 256]$, $\nu = 5$ and Chebyshev filter size to be 3, and train 100 epochs with initial learning rate of 0.01 via Adam optimizer (Kingma & Ba, 2015). All experiments are bootstrapped with 10 runs using random seeds.

For graph regression tasks, we generally vary the graph types, model depths, and max orders, as shown in Figure 1, 5, and 6.

B.2. Details on Synthesized Heterophilic Graphs and Missing Experimental Details

We generate graphs with heterophily that often have contrasting node labels between neighbors. We synthesize such graphs by firstly generating random node labels according to the number of classes, then making edges between nodes with different labels at a higher probability p_{hetero} (e.g. 0.8).

We learn from Theorem 5.2 that many-body MPNN may generate higher energies than ChebNet, which makes it less prone from over-smoothing issues. It is notable that while generating much energy, it does not suffer from over-fitting.

To isolate the contribution of each interaction term of many-body MPNN, we plot the logarithmic energy growth for each k-body component, averaged by layers (Figure 4). We observe that the energy contribution is increasing over training process, while higher-body term contributes consistently higher energies as stated in Lemma A.1.

B.3. Evaluating Mixing on Synthetic Spine Graphs

To test the sensitivity bounds, we synthesized spine graphs with each node on the spine having multiple one-hop neighbor nodes attached only to it. If any edge on spine is erased, the graph becomes disconnected; hence these edges have generally low Ricci curvatures.

To evaluate the mixing power of many-body interactions, we identify node groups of sizes [3 to ν] and derive the Jacobians $\max(\partial \boldsymbol{h}_{\boldsymbol{u}}^{(\boldsymbol{r})} / \partial \boldsymbol{x}_v)$, with the maximum taken across embedding dimensions. We average the Jacobians from each group to represent the mixing power. We open-source the notebook for future exploration.

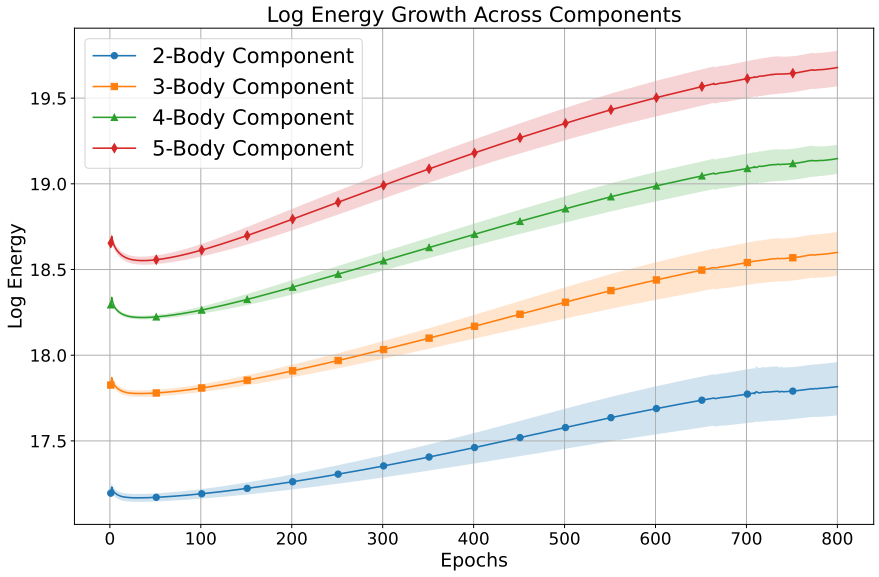

Figure 4. Log-energy growth from k-body term component ($k \leq 5$) of Many-body MPNN.

B.4. On Regressing Synthetic Graph Energies

B.4.1. Fixed-Shaped Graphs

We also experiment on regular-shaped graphs such as `Ring`, `CrossedRing`, `CliquePath` as with Topping et al. (2022), and `Spine` graphs having a long spine with each attached to a fixed number of one-hop neighbors. We synthesize graph energies in an alternative manner, based on their spectral properties: we take the absolute sum of eigenvalues of graph distance matrix that contains the pairwise shortest path lengths, or that of the graph adjacency matrix A, for emphasizing either node distances or local clustering.

The intuition for such energy construction is that graphs with more connectivity have wider spectral gaps, leading to larger target values, whilst making the formulation non-linear and not be modeled precisely with simple two-body message passing.

Our many-body formulation models quite precisely the spectral properties of these regular-shaped graphs (Figure 5), and runs inductively and faster than eigen-decomposition on larger graphs due to lower time complexity (Proposition 5.3). We observe that many-body MPNN often achieves the best performance when more layers (e.g., ≥ 32) are stacked, indicating that it is less prone to over-squashing issues.

B.4.2. On Correlation Orders

We study different correlation orders and extract the patterns of performances on fixed-shaped graphs (see Figure 6). We observe that generally across graph shapes, more layers lead to more fluctuations when increasing max orders, and more layers with high orders will likely lead to over-smoothing. We leave the detailed investigation of effective correlation orders with graph topology for future work.

C. Compute Resources

We utilize one slice of A100 GPUs 80GB card (and one entire NVIDIA RTX 2080 Ti only for benchmarking) for speeding up graph convolutions and fully-connected layer operations, and AMD CPUs for graph synthesis and energy generation.

The experiments involved working with graph convolutional models on synthesized graphs of approximately 10,000 nodes with edge probabilities ranging from 0.15 to 0.3. The models consisted of 1-32 convolutional layers plus 2 linear layers, using the Adam optimizer, and with hidden dimensions ranging from 2 to 256. There are 5 types of graphs each with 100 instances, 4 types of models, and 10 repeated experiments for each setting.

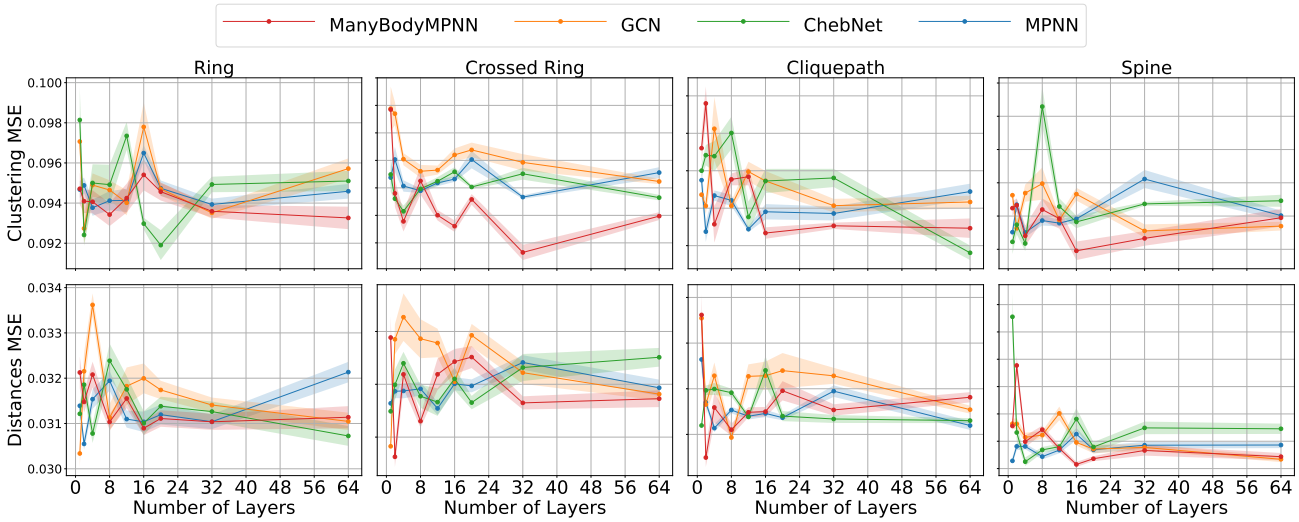

Figure 5. Test MSE on regressing the synthetic energies of regular-shaped graphs when varying the number of layers from 1 to 32 for different models. The graph energy functions are different between emphases on node distances and clustering and hence the two rows of plots only share the x-axis.

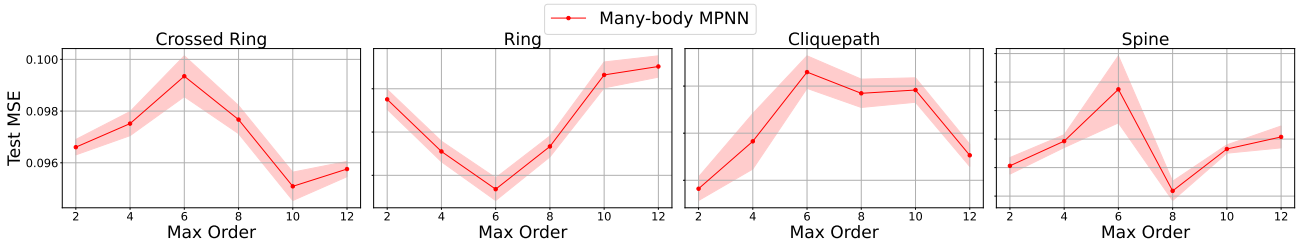

Figure 6. Test MSE for many-body MPNN of particular max correlation orders for regressing synthetic graph energies. There are 100 synthetic graphs of 300-500 nodes. The energy function is clustering-emphasized.

Based on the rough estimate of running 2 hours for each experiment series (e.g. B.4.1; there are 3 such series), we estimate $2 \times 4 \times 3 \times 10 = 240$ GPU-hours to calculate the results in the paper, plus around 300 CPU-hours.