Framelet Message Passing: GNNs Propagation with Multiscale Multi-hop Representation and No Oversmoothing

Xinliang Liu
KAUST
xinliang.liu@kaust.edu.sa

Bingxin Zhou
University of Sydney
bzho3923@uni.sydney.edu.au

Chutian Zhang
Shanghai Jiao Tong University
scarborough@sjtu.edu.cn

Yu Guang Wang†
Shanghai Jiao Tong University
yuguang.wang@sjtu.edu.cn

***Abstract*—Graph neural networks (GNNs) encode adequate feature representations of graphs. The dominant graph feature distillation leverages neural message passing which is a message update method of a specific node feature from one layer to the next by incorporating 1-hop neighbors. It nevertheless usually suffers from oversmoothing. This work proposes Framelet Message Passing that integrates multiscale framelet representation of neighbor nodes in multiple hops in node message update. Our method circumvents oversmoothing with a non-decay Dirichlet energy in propagation. It achieves state-of-the-art performance on real node classification tasks with low computational cost.**

I. Introduction

Graph neural networks (GNNs) have received growing attention in past five years [1], [2], [3]. GNNs equip effective graph convolutions to distill useful features and structural information of given graph signals. Existing methods usually summarize a node's local properties from its spatially-connected neighbors. Such a scheme is called message passing (MP) [4], where different methods differentiate each other by their unique design of the aggregator. Most of these convolutions are based on the first-order approximation of eigendecomposition by the graph adjacency matrix, and they are proved recklessly removing high-pass information [5], [6], [7]. Consequently, many local details are lost during forward propagation. This deficiency limits the expressivity of GNNs and partly gives rise to the oversmoothing issue of a deep GNN. A few existing spectral-involved MP schemes [8], [9], [10] use eigenvectors to feed projected node feature into the aggregator. They capture the directional flow in the input by Fourier transforms but overlook the power of multi-scale representation. This work establishes an MP scheme using mulitscale graph framelet transforms. Framelet low and high decomposition for a graph can be employed to represent neighborhood message for the feature update of an individual node in MP. The message update aggregates a node's last input feature with its scale-wise framelet representation. The latter is a weighted amalgamation of the node's multi-hop neighbor features in the framelet domain, where the weights are from the graph rewiring of framelet coefficients.

II. Main Results

An undirected attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{X})$ consists of a non-empty finite set of $n = |\mathcal{V}|$ nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$ between node pairs. Denote $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ the (weighted) graph adjacency matrix and $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ the node attributes. A graph convolution learns a matrix representation $\boldsymbol{H}$ that embeds the structure $\boldsymbol{A}$ and feature matrix $\boldsymbol{X} = \{\boldsymbol{X}_j\}_{j=1}^{N}$ with $\boldsymbol{X}_j$ for node j. Most graph convolutions follow the MP [4] scheme, which finds a central node's smooth representation by aggregating its 1-hop neighbor information. At a specific layer, the propagation for the ith node reads

$$\boldsymbol{X}_i^{\rm out} = \gamma\left(\boldsymbol{X}_i^{\rm in}, \square_{j\in\mathcal{N}(i)}\phi(\boldsymbol{X}_i^{\rm in}, \boldsymbol{X}_j^{\rm in}, \boldsymbol{A}_{ij})\right),$$

where $\square$ is a differentiable and permutation invariant aggregation function. The set $\mathcal{N}(i)$ includes $\mathcal{V}_i$ and its 1-hop neighbors.

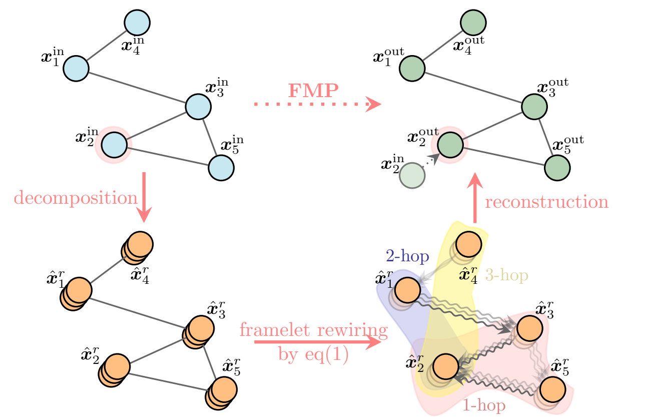

Fig. 1. An illustrative workflow of framelet message passing that propagates $\boldsymbol{X}^{\rm out}$ (green) from $\boldsymbol{X}^{\rm in}$ (blue). An input graph signal is first decomposed into multi-scale coefficients (orange) in the framelet domain. A coefficient $\hat{\boldsymbol{x}}_2$ at the rth pass aggregates multi-hop neighbor messages with graph rewiring. The framelet representations are then projected back to establish $\boldsymbol{x}_2^{\rm out}$.

In a spectral-based graph convolution, graph signals are transformed to a set of coefficients $\hat{\boldsymbol{X}} = \mathcal{W}\boldsymbol{X}^{\rm in}$ in frequency channels to learn useful graph representation. With K high-pass filters, $\hat{\boldsymbol{X}} = \{\mathcal{W}_{0,0}\boldsymbol{X}; \mathcal{W}_{r,l}\boldsymbol{X} : r = 1,\ldots,K, l = 1,\ldots,J\}$ constitutes of a low-pass coefficient matrix $\mathcal{W}_{0,0}\boldsymbol{X}$ and K high-pass coefficient matrices $\mathcal{W}_{r,l}\boldsymbol{X}$ at J scale levels. The framelet coefficient at node i collects information from its neighbors in the framelet domain of the same channel. As the original (undecimated) framelet transforms are integrals, we need Chebyshev polynomial approximation for $\mathcal{W}_{r,l}$ (including $\mathcal{W}_{0,0}$), and then obtain the corresponding *approximated framelet transform matrices* $\mathcal{W}_{r,l}^{\natural}$ as products of Chebyshev polynomials of $\boldsymbol{A}$ [11], [12]. Let m be the highest order of the Chebyshev polynomial involved. The aggregation of framelet coefficients for the neighborhood of the node i takes over up to the m-th multi-hop $\mathcal{N}^m(i)$. **Framelet message passing** (FMP) is defined by, for the update of graph node feature $\boldsymbol{X}^{(t)}$ at layer t,

$$\boldsymbol{X}^{(t+1)} = \boldsymbol{X}^{(t)} + \boldsymbol{Z} \tag{1}$$

$$\boldsymbol{Z}_i^r = \gamma\left(\sum_{j\in\mathcal{N}^m(i)}\left(\sum_{r=1}^{K}\sum_{l=1}^{J}(\theta_r\mathcal{W}_{r,l}^{\natural})_{i,j} + (\theta_0\mathcal{W}_{0,0}^{\natural})_{i,j}\right)\boldsymbol{X}_j^{(t)}\right),$$

where $\boldsymbol{Z}_i^r$ is the propagated feature at node i over the low and high passes and all scales. Aggregating the framelet coefficients from an m-order approximated $\mathcal{W}_{r,l}^{\natural}$ is as powerful as conducting an m-hop neighborhood spatial MPNN [4]. Apart from this, we can use a continuous update scheme with ODE for FMP in (1):

$$\mathrm{d}\boldsymbol{X}_i(t)/\mathrm{d}t = \boldsymbol{Z}_i^r. \tag{2}$$

Here $\boldsymbol{X}_i(t)$ is the feature of node i at continous time t. The equation (2) can be effectively solved by a numerical method like

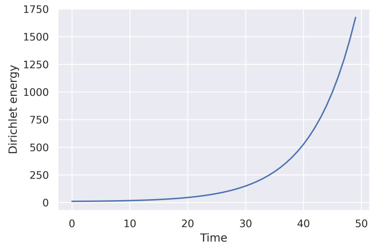

Fig. 2. Energy evolution for $\mathrm{FMP}_{\mathrm{ode}}$. The energy increases over network propagation. The oversmoothing issue in GNNs is circumvented with $\mathrm{FMP}_{\mathrm{ode}}$.

Dormand–Prince adaptive step size scheme [13] with a better feature extraction performance.

Oversmoothing is a serious problem for most deep GNN models. For example, graph convolutional network (GCN) [14] has the Dirichlet energy decaying rapidly to zero as the network depth increases [15]. The Dirichlet energy of a node feature $\boldsymbol{X}$ from $\mathcal{G}$ with graph Laplacian $\mathcal{L} = I - \boldsymbol{A}$ is defined by

$$E(X) = \mathrm{tr}(X^\top \widetilde{\mathcal{L}} X) = \frac{1}{2}\sum \boldsymbol{A}_{ij}\left(\frac{\boldsymbol{X}_i}{\sqrt{1+d_i}} - \frac{\boldsymbol{X}_j}{\sqrt{1+d_j}}\right)^2,$$

where $\widetilde{\mathcal{L}} := D^{-1/2}\mathcal{L}D^{-1/2}$ is the normalized graph Laplacian with degree $D = \mathrm{diag}(d_1, \ldots, d_N)$, $d_j = \sum_{i=1}^N \boldsymbol{A}_{ij}$. Our FMP provides a remedy for the oversmoothing predicament by decomposing $\boldsymbol{X}$ into low pass and high passes with the energy conservation property

$$E(\boldsymbol{X}) = E(\mathcal{W}_{0,0}\boldsymbol{X}) + \sum_{r=1}^{K}\sum_{l=1}^{J} E(\mathcal{W}_{r,l}\boldsymbol{X}), \tag{3}$$

where $E(\mathcal{W}_{0,0}\boldsymbol{X})$ and $E(\mathcal{W}_{r,l}\boldsymbol{X})$ break down the total energy $E(\boldsymbol{X})$ into multi scales, and low and high passes. Note that we use the original framelet transforms here. For the approximated framelet transforms $\mathcal{W}^{\natural}_{r,l}$, there is an energy change on RHS of (3) due to the truncation error. The relatively smoother node features account for a lower level energy $E(\mathcal{W}_{0,0}\boldsymbol{X})$. In this way, we provide more flexibility to maneuver energy evolution as the energy does not decay when we use FMP in (1) or (2). Under some mild assumptions on γ and θ_r, we have the following two estimates on energy.

Theorem 2.1: The Dirichlet energy of the output of a framelet message passing (FMP) layer by (1) is larger than twice the input's.

$$E(\boldsymbol{X}^{\mathrm{out}}) \geq 2E(\boldsymbol{X}^{\mathrm{in}}).$$

Theorem 2.2: The framelet message passing with ODE update scheme ($\mathrm{FMP}_{\mathrm{ode}}$) in (2) has the Dirichlet energy

$$E(\boldsymbol{X}(0)) \leq E(\boldsymbol{X}(t)) \leq e^{\sqrt{J}t}E(\boldsymbol{X}(0)).$$

We validate the performance of FMP with node classification on three citation networks [16]. The mean accuracy is reported in Table I along with baseline methods which results are retrieved from [17]. Our model uses a 2-layer GNN with FMP. Two types of FMP as defined above are considered: $\mathrm{FMP}_{\mathrm{mlp}}$ in (1) with γ a 2-layer MLP, and $\mathrm{FMP}_{\mathrm{ode}}$ in (2) with a linear aggregator. Our methods achieve the top performance on all three tasks. This shows that FMP has better learning ability as multiple hops' information is taken into account.

TABLE I

AVERAGE ACCURACY OF NODE CLASSIFICATION OVER 10 REPETITIONS.

	Cora	**Citeseer**	**Pubmed**
MLP	57.8±0.11	61.2±0.08	73.2±0.05
ChebNet [18]	79.9±0.11	50.8±0.00	72.7±0.00
GCN [14]	82.4±0.25	70.7±0.36	79.4±0.15
GAT [19]	82.5±0.31	72.1±0.41	79.1±0.22
GraphSAGE [20]	82.1±0.25	71.8±0.36	79.2±0.27
SGC [5]	81.9±0.23	72.2±0.22	78.3±0.14
JKNet [21]	81.2±0.49	70.7±0.88	78.6±0.25
$\mathrm{FMP}_{\mathrm{mlp}}$ (ours)	82.2±0.70	71.7±0.40	79.4±0.30
$\mathrm{FMP}_{\mathrm{ode}}$ (ours)	**83.6**±0.80	**72.7**±0.90	**79.6**±0.90

III. CONCLUSION

This work proposes an expressive framelet message passing for GNNs propagation. The framelet coefficients of neighboring nodes provide a graph rewiring scheme to amalgamate features in the framelet domain. We show that our FMP circumvents oversmoothing that appears in most spatial GNN methods. The spectral information reserves extra expressivity to the graph representation by taking multiscale framelet representation into account. The proposed framelet message passing has low computational complexity.

REFERENCES

[1] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond Euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[2] W. L. Hamilton, "Graph representation learning," *Synthesis Lectures on Artifical Intelligence and Machine Learning*, vol. 14, no. 3, pp. 1–159, 2020.

[3] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE TNNLS*, vol. 32, no. 1, pp. 4–24, 2020.

[4] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *ICML*, 2017.

[5] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *ICML*, 2019.

[6] K. Oono and T. Suzuki, "Graph neural networks exponentially lose expressive power for node classification," in *ICLR*, 2019.

[7] D. Bo, X. Wang, C. Shi, and H. Shen, "Beyond low-frequency information in graph convolutional networks," in *AAAI*, vol. 35, no. 5, 2021, pp. 3950–3957.

[8] K. Stachenfeld, J. Godwin, and P. Battaglia, "Graph networks with spectral message passing," in *NeurIPS 2020 Workshop on Interpretable Inductive Biases and Physically Structured Learning*, 2020.

[9] M. Balcilar, P. Héroux, B. Gauzere, P. Vasseur, S. Adam, and P. Honeine, "Breaking the limits of message passing graph neural networks," in *ICML*, 2021, pp. 599–608.

[10] D. Beaini, S. Passaro, V. Létourneau, W. Hamilton, G. Corso, and P. Liò, "Directional graph networks," in *ICML*, 2021, pp. 748–758.

[11] B. Dong, "Sparse representation on graphs by tight wavelet frames and applications," *Applied and Computational Harmonic Analysis*, vol. 42, no. 3, pp. 452–479, 2017.

[12] X. Zheng, B. Zhou, J. Gao, Y. G. Wang, P. Lio, M. Li, and G. Montúfar, "How framelets enhance graph neural networks," in *ICML*, 2021.

[13] J. R. Dormand and P. J. Prince, "A family of embedded Runge-Kutta formulae," *Journal of Computational and Applied Mathematics*, vol. 6, no. 1, pp. 19–26, 1980.

[14] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017.

[15] C. Cai and Y. Wang, "A note on over-smoothing for graph neural networks," *ICML Workshop on Graph Representation Learning*, 2020.

[16] Z. Yang, W. Cohen, and R. Salakhudinov, "Revisiting semi-supervised learning with graph embeddings," in *ICML*, 2016.

[17] M. Zhu, X. Wang, C. Shi, H. Ji, and P. Cui, "Interpreting and unifying graph neural networks with an optimization framework," in *Proceedings of the Web Conference*, 2021, pp. 1215–1226.

[18] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *NIPS*, 2016.

[19] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *ICLR*, 2018.

[20] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *NIPS*, 2017.

[21] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, and S. Jegelka, "Representation learning on graphs with jumping knowledge networks," in *ICML*, 2018.