
How Framelets Enhance Graph Neural Networks

Xuebin Zheng^{*1} Bingxin Zhou^{*1} Junbin Gao¹ Yu Guang Wang²³⁴
Pietro Liò⁵ Ming Li⁶ Guido Montúfar⁷²

Abstract

This paper presents a new approach for assembling graph neural networks based on framelet transforms. The latter provides a multi-scale representation for graph-structured data. We decompose an input graph into low-pass and high-pass frequencies coefficients for network training, which then defines a framelet-based graph convolution. The framelet decomposition naturally induces a graph pooling strategy by aggregating the graph feature into low-pass and high-pass spectra, which considers both the feature values and geometry of the graph data and conserves the total information. The graph neural networks with the proposed framelet convolution and pooling achieve state-of-the-art performance in many node and graph prediction tasks. Moreover, we propose shrinkage as a new activation for the framelet convolution, which thresholds high-frequency information at different scales. Compared to ReLU, shrinkage activation improves model performance on denoising and signal compression: noises in both node and structure can be significantly reduced by accurately cutting off the high-pass coefficients from framelet decomposition, and the signal can be compressed to less than half its original size with well-preserved prediction performance.

1. Introduction

Graph neural networks (GNNs) are a powerful deep learning method for prediction tasks on graph-structured data (Wu et al., 2021). Most existing GNN models are spatial-based, such as GCN (Kipf & Welling, 2017), GAT (Veličković et al., 2018) and GIN (Xu et al., 2019b). These methods compute graph convolution on vertices and edges in the form of message passing (Gilmer et al., 2017), but leave the signal frequency of graph data unexplored. In this paper, we seek to exploit signal processing for GNNs. Graph framelets (Dong, 2017; Zheng et al., 2020b), akin to traditional wavelets, provide a multiresolution analysis (MRA) for graph signals. The fully tensorized *framelet transforms* guarantee an efficient design of graph convolution that combines low-pass and high-pass information, where the transforms only require graph Laplacian, Chebyshev polynomial approximation, and filter banks. We propose *framelet convolution* that exploits the decomposition and reconstruction procedures of framelet transforms, and the network learns in the frequency domain.

The wavelet-like graph data analysis allows us to exploit traditional tools from signal processing. An effective practice is the *shrinkage* that thresholds high-pass coefficients in the framelet representation. The multi-scale property of framelet convolution introduces diffusion for GNNs. The associated shrinkage defines a new type of activation that adapts to the diffusion scales in nonlinear feature transformation and extraction. Shrinkage in framelet convolution provides a mechanism for effective noise reduction of input graph data, where noise may appear in node and/or edge features. This ability is inherited from the traditional wavelet denoising model. Moreover, the shrinkage activation provides a way to compress graph signals. The shrinkage trims coefficients and significantly compresses the graph data representation while at the same time the underlying GNN reaches state-of-the-art performance in multiple tasks. The framelet MRA and shrinkage threshold provide GNNs with multi-scale and compression characteristics, which distinguishes our model from existing graph convolution methods.

The framelet transform naturally induces a graph pooling strategy by aggregating the different scales of framelet features. The consequent framelet pooling conserves the

^{*}Equal contribution ¹The University of Sydney Business School, The University of Sydney, Camperdown, NSW 2006, Australia. ²Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany. ³Institute of Natural Sciences and School of Mathematical Sciences, Shanghai Jiao Tong University, China. ⁴School of Mathematics and Statistics, The University of New South Wales, Sydney, Australia. ⁵Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom. ⁶Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University, Jinhua, China. ⁷Department of Mathematics and Department of Statistics, University of California, Los Angeles, United States. Correspondence to: Bingxin Zhou <bzho3923@uni.sydney.edu.au>, Yu Guang Wang <yuguang.wang@mis.mpg.de>.

total information due to energy conservation of framelet spectra, and offers an efficient graph dimensionality reduction. Framelet-based GNNs outperform existing spatial or spectral-based methods on benchmark node and graph prediction tasks. Moreover, the framelet convolution with ReLU or shrinkage can achieve excellent performance in denoising both node features and graph structure. This behavior suggests that the framelets play an important role in bridging signal processing and graph neural networks.

2. Related Works

Graph Framelets and Transforms The construction of wavelet analysis on graphs was first explored by Crovella & Kolaczyk (2003). Maggioni & Mhaskar (2008) used polynomials of a differential operator to build multi-scale transforms. The spectral graph wavelet transforms (Hammond et al., 2011) define the graph spectrum from a graph’s Laplacian matrix, where the scaling function is approached by the Chebyshev approximation. Behjat et al. (2016) encoded energy spectral density to design tight frames on graphs with both graph topology and signal features. Dong (2017) approximated piece-wise smooth functions with undecimated tight wavelet frames. Fast decomposition and reconstruction become possible with the filtered Chebyshev polynomial approximation and proper design of filter banks.

The other regime of signal processing on graph data is backboneed by the multiresolution analysis (Mallat, 1989) that establishes a tree-based wavelet system with localization properties. Crovella & Kolaczyk (2003) defined the ‘h-hop’ neighborhoods on binary graphs and Gavish et al. (2010) constructed Haar-like bases. The Haar-like orthonormal wavelet system (Chui et al., 2015) has been applied to deep learning tasks on undirected graphs (Wang et al., 2020b; Li et al., 2020b; Zheng et al., 2020a). Meanwhile, fast tight framelet filter bank transforms on quadrature-based framelets are explored on graph domain (Wang & Zhuang, 2019; Zheng et al., 2020b) and manifold space (Wang & Zhuang, 2018) with a low redundancy rate.

Graph Convolution and Graph Pooling The theory of graph convolution (Bruna et al., 2014) facilitated the later development of advanced deep learning methods. For example, spectral-based GNNs (Defferrard et al., 2016a; Xu et al., 2019a; Li et al., 2020b; Balcilar et al., 2021) transform graph signals to the spectral domain and process them with filter operations. Alternatively, spatial-based graph convolution performs node property prediction via aggregating feature information over neighborhood nodes (Kipf & Welling, 2017; Gilmer et al., 2017; Wang et al., 2020a; Vignac et al., 2020; Chen et al., 2020). For graph property prediction, one pursues topology-aware graph embedding via graph pooling. Some global pooling strategies refine

vertex features in one-shot (Zhang et al., 2018b; Lee et al., 2019), while others process graph information hierarchically (Cangea et al., 2018; Gao & Ji, 2019; Knyazev et al., 2019; Wang et al., 2020b; Ma et al., 2020).

Signal Compression and Denoising Signal compression is critical for high-speed signal transmission. Wavelets play an important role in compressing signal and have contributed to the prevalent JPEG 2000 (ISO, 2019). Our shrinkage framelet convolution provides an algorithm for compressing graph signals. Denoising problems have long been studied in image processing. Many models have been proposed for image restoration (Milanfar, 2013). In particular, wavelets provide a sparse and multi-scale representation for images and have proved an impressive regularizer for reducing Gaussian white noise for signals in 2D (Figueiredo & Nowak, 2003; Cai et al., 2012; Dong & Shen, 2013; Shen, 2010). Graph spectral theory and graph wavelets have been widely used for image processing (Cheung et al., 2018). Our convolution uses graph framelets and provides a solution to the denoising model for structured data using GNN training.

3. Multiresolution Analysis of Graph Framelets

Our convolution uses the undecimated graph framelets and their transforms, which were introduced by Dong (2017, Section 3); Zheng et al. (2020b, Section 4.1). Framelets on a specific graph $\mathcal{G} = (V, E, \omega) \in L_2(\mathcal{G})$ are defined by a *filter bank* and the spectrum of its graph Laplacian \mathcal{L} . The filter bank $\boldsymbol{\eta} := \{a; b^{(1)}, \dots, b^{(n)}\} \in l_0(\mathbb{Z})$ for a framelet system is a set of compactly supported sequences, where n denotes the number of high pass filters. The *low-pass* and *high-pass filters* of the framelet transforms, a and $b^{(r)}$, distill and represent the approximation and detail information of the graph signal. The scaling functions $\Psi = \{\alpha; \beta^{(1)}, \dots, \beta^{(n)}\}$ with respect to the filter bank $\boldsymbol{\eta}$ are used to generate the framelets. The $\alpha, \beta^{(r)}$ and their Fourier transforms are in $L_2(\mathbb{R})$. For $\xi \in \mathbb{R}$, the filters satisfy the classic refining equations

$$\widehat{\alpha}(2\xi) = \widehat{a}(\xi)\widehat{\alpha}(\xi), \quad \widehat{\beta^{(r)}}(2\xi) = \widehat{b^{(r)}}(\xi)\widehat{\alpha}(\xi), \quad r = 1, \dots, n.$$

Graph Framelets Suppose $\{(\lambda_\ell, \mathbf{u}_\ell)\}_{\ell=1}^N$ are the eigenvalue and eigenvector pairs for \mathcal{L} of graph \mathcal{G} with N nodes. The (undecimated) framelets at *scale level* $j = 1, \dots, J$ for graph \mathcal{G} with the above scaling functions are defined, for $n = 1, \dots, r$, by

$$\begin{aligned} \varphi_{j,p}(v) &= \sum_{\ell=1}^N \widehat{\alpha} \left(\frac{\lambda_\ell}{2^j} \right) \overline{\mathbf{u}_\ell(p)} \mathbf{u}_\ell(v) \\ \psi_{j,p}^n(v) &= \sum_{\ell=1}^N \widehat{\beta^{(n)}} \left(\frac{\lambda_\ell}{2^j} \right) \overline{\mathbf{u}_\ell(p)} \mathbf{u}_\ell(v), \end{aligned} \tag{1}$$

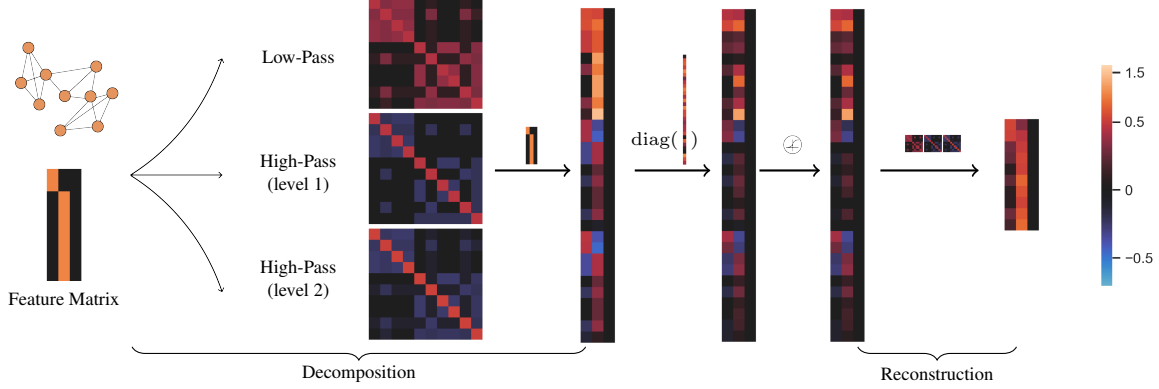


Figure 1. Computational flow of the proposed undecimated framelet graph convolution (UFGConv). This is an illustration with shrinkage activation, which will be discussed in detail in Section 4. Given a graph with structure (adjacency matrix) and feature information, the target is to properly embed the graph by graph convolution. The demonstrative sample is a graph with 10 nodes and 3 features extracted from **PROTEINS** in **TUDataset**. The framelet dilation and scale level are both set to default value 2. The UFGCONV applies tensor-based framelet transform, and constructs one low-pass and two high-pass *framelet transform matrices* $\mathcal{W}_{r,j}$, which are then multiplied by the input feature matrix to produce the framelet coefficients. Moreover, these coefficients are processed by the trainable network filter and compressed by the shrinkage. Finally, the activated coefficients are reconstructed and sent back to the spatial domain as the convolution output by using the framelet transform matrices again with transposed alignment.

where $\varphi_{j,p}$ or $\psi_{j,p}^r$ is the low-pass or high-pass framelet translated at node p . The low-pass and high-pass *framelet coefficients* for a signal f on graph \mathcal{G} are $v_{j,p}$ and $w_{j,p}^r$, which are the projections $\langle \varphi_{j,p}, f \rangle$ and $\langle \psi_{j,p}^r, f \rangle$ of the graph signal onto framelets at scale j and node p . Here we use Haar-type filters for framelets (Dong, 2017). The dilation factor is 2^j with the *dilation* (base) 2.

The above framelet system is *tight* if it provides an exact representation for any function in $L_2(\mathcal{G})$. The tightness is determined by filter banks (See Theorem 1 in Appendix B). This condition guarantees a unique representation of a graph signal with framelet coefficients. It also helps manipulate the graph data in the framelet (frequency) domain.

The graph convolution developed in this work with a tight graph framelet system is effective for several reasons. First, the undecimated framelet transforms have a simplified formula that uses only graph Laplacian and filtered Chebyshev polynomial approximation, as will be further explained below. The resulting framelet transforms can be written as a sparse tensor and the corresponding framelet coefficients are evaluated fast in tensor computation. See Figure 1 for the computational flow of framelet convolution. Moreover, the tight framelets on the graph have a low redundancy rate, which is analogous to the framelets on a manifold as considered by Wang & Zhuang (2018; 2019). We give more discussion about framelets in the Appendix.

Tensor-based \mathcal{G} -framelet Transforms Framelet transforms map between a graph signal f and its representative framelet coefficients. We point out an approximate formula for (1), which exploits the Chebyshev polynomial approximation to the filters $\widehat{\alpha}$ and $\widehat{\beta}^{(r)}$, and thus gives a

simplified version and fast evaluation for framelet transforms. By the framelet transform theorem (in Appendix), the corresponding framelet transforms are implemented by \mathcal{W} and \mathcal{V} , the decomposition and reconstruction operators. For graph signal f , we define $\mathcal{W} = \{\mathcal{W}_{r,j} | r = 1, \dots, n; j = 1, \dots, J\} \cup \{\mathcal{W}_{0,J}\}$ entry-wisely, where $\mathcal{W}_{r,1}f = \mathcal{T}_r^k(2^{-J}\mathcal{L})f$, and

$$\mathcal{W}_{r,j}f = \mathcal{T}_r^k(2^{K+j-1}\mathcal{L})\mathcal{T}_0^k(2^{K+j-2}\mathcal{L})\dots\mathcal{T}_0^k(2^{-K}\mathcal{L})f$$

for $j = 2, \dots, J$. Here \mathcal{T}_r^k is the r -degree Chebyshev polynomial, \mathcal{L} is the graph Laplacian, and K is the constant determined by the maximum eigenvalue of \mathcal{L} that satisfies $\lambda_{\max} \leq 2^K\pi$. The $\mathcal{W}_{0,J}f = \{v_{J,p}\}_{p \in V}$ are the *low-pass coefficients* and $\mathcal{W}_{r,j}f = \{w_{j,p}^r\}_{p \in V}$ are *high-pass coefficients* of f . The j indicates the scale level, and $r = 1, \dots, n$ with n the number of high-passes. The reconstruction operator \mathcal{V} is the realignment of the framelet transform matrices of the decomposition operator \mathcal{W} .

Figure 1 gives the fast \mathcal{G} -framelet transform algorithm based on tensorized \mathcal{W} and \mathcal{V} with scale level 2 and shrinkage threshold $\sigma = 1$ (the meaning of σ will be discussed in detail in Section 5). In practice, we turn the computation into merely sparse matrix multiplication by properly aligning the low-pass and high-pass elements of \mathcal{W} and \mathcal{V} . The tensor-based framelet transforms have time complexity $\mathcal{O}(N^2(nJ+1)Kd)$ and space complexity $\mathcal{O}(N^2(nJ+1)d)$ for an N -node graph and d features. Here, the n , J and K are constants independent of graph data. See Appendix for an empirical study of the complexity of framelet transforms on benchmarks.

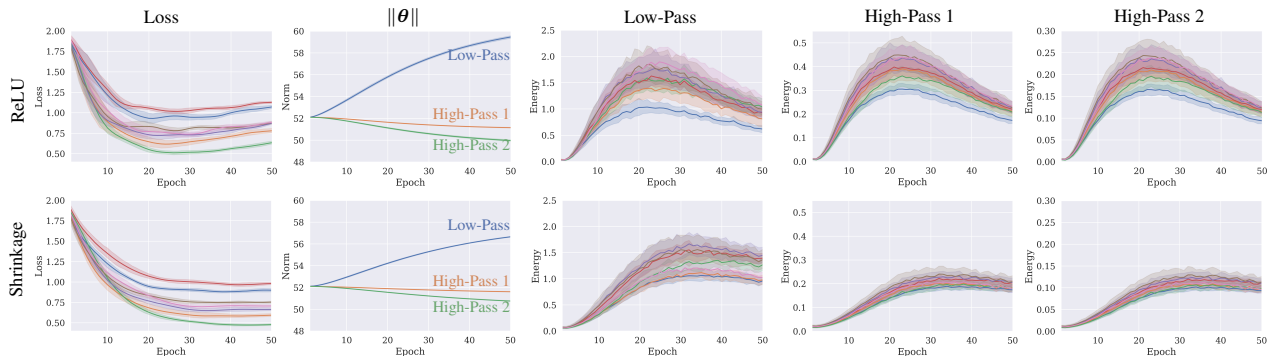


Figure 2. Learning behavior of the final framelet convolutional layer of GNN with two UFGCONV for **Cora**. Top is for UFGCONV with ReLU activation in (2). Bottom is for UFGCONV with Shrinkage activation in (3). From left to right we show some key network learning properties: training loss, l_2 norm of network filter θ , power spectrum of framelet coefficients for low-pass and high-passes at scale levels 1 and 2. The curves of the quantities for each of 7 label classes are shown. Framelets provide a good feature representation and shrinkage makes training more stable, where central information energy is conserved via thresholding high-pass coefficients.

4. Framelet Convolution

With the above \mathcal{G} -framelet transforms, we can define a *framelet (graph) convolution* similar to the spectral graph convolution. For *network filter* θ and input *graph feature* $X \in \mathbb{R}^{N \times d}$ of the graph \mathcal{G} with N nodes, we define

$$\theta \star X = \text{ReLU}(\mathcal{V}(\text{diag}(\theta)(\mathcal{W}X'))), \quad X' = XW. \quad (2)$$

As mentioned, $\mathcal{W}X'$ is the framelet coefficient matrix for the transformed X' , where \mathcal{W} is a sequence of $nJ + 1$ transform matrices (each of size $N \times N$) for low-pass and high-passes. The size of the vector θ is $(nJ + 1)N$ which matches the total number of the framelet coefficients for each feature. The network filter θ lies in the frequency domain, each component of which is multiplied to the corresponding row of $\mathcal{W}X'$. The matrix W in (2) is a trainable weight matrix with dimension $d \times d'$.

We can replace the ReLU activation in (2) with a wavelet shrinkage threshold function, or *shrinkage*. As the framelet coefficients lie in multiple scales, the “one size fits all” criterion of ReLU could potentially damage the multi-scale property of the framelet representation. In contrast, the shrinkage threshold adapts to the varying scales of the coefficients and provides a more precise cutoff. Thus we can use fewer coefficients while maintaining a comparable performance in framelet representation, see Section 8.1. The framelet convolution with *shrinkage activation* is

$$\mathcal{V}(\text{Shrinkage}(\text{diag}(\theta)(\mathcal{W}X'))), \quad X' = XW. \quad (3)$$

Different from ReLU which works in the spatial domain, in (3), the shrinkage activation is carried on before the framelet reconstruction \mathcal{V} as the shrinkage thresholds for the high-pass coefficients in the framelet domain. Figure 1 demonstrates a computational flow of the framelet convolution that learns the graph embedding of a graph with feature matrix $X^{\text{in}} \in \mathbb{R}^{N \times d}$ by a framelet system of 2 scale levels ($j = 1, 2$) and 1 high-pass filter ($r = 0, 1$). Here, we

omit the feature transformation step for a simple illustration. The decomposition $\mathcal{W} = \{\mathcal{W}_{0,2}; \mathcal{W}_{1,1}, \mathcal{W}_{1,2}\}$ transforms the input feature matrix with one low-pass and two high-pass operators. The operators can be rearranged to $\mathcal{W}^{\natural} := [\mathcal{W}_{0,2}, \mathcal{W}_{1,1}, \mathcal{W}_{1,2}]^{\top}$ of three $N \times N$ sparse matrices. The coefficients $\mathcal{W}^{\natural}X^{\text{in}} \in \mathbb{R}^{3N \times d}$ can then be obtained by matrix multiplication. The network learning propagates in the frequency domain, where one applies the network filter θ to the framelet coefficients $\mathcal{W}^{\natural}X^{\text{in}}$. For the framelet convolution with shrinkage, the filtered coefficients would be activated before applying $\mathcal{V}^{\natural} = (\mathcal{W}^{\natural})^{\top}$. Otherwise the ReLU activation will act on the reconstructed signal in the spatial domain.

5. Shrinkage and Signal Compression

Wavelet shrinkage is intimately linked to multiresolution properties of the wavelet transform in the classic wavelet theory. The shrinkage only applies to finer scales, i.e., detail coefficients (Donoho et al., 1995), so that the wavelet scalogram (a paradigm of the time-frequency energy localization of a signal) experiences a minimal change. This property promises a meaningful estimator for signal compression and an explainable graph convolution with framelets.

Sparsity and Compression The high-pass coefficients in the frequency domain can be cut off by shrinkage thresholding. For example, the *soft-thresholding* (Donoho & Johnstone, 1994; Donoho et al., 1995; Tibshirani, 1996) defines

$$\text{Shrinkage}(x) = \text{sgn}(x)(|x| - \lambda)_+ \quad \forall x \in \mathbb{R},$$

where λ is the threshold value. Any x with its absolute value less than λ shall return to zero. Applying the above soft-thresholding to the shrinkage activation in (3) only influences small high-pass framelet coefficients. We also consider the scale-dependent selection threshold with demarcation point (Donoho, 1995): $\lambda = \sigma \sqrt{2 \log(N)} / \sqrt{N}$

for N coefficients. The hyperparameter σ is an analogue to the noise level of the wavelet denoising model. We let σ be associated with the magnitude order of the coefficients so it reflects the scale of the framelet representation.

The shrinkage benefits framelet graph convolution by reducing noise in framelet coefficients and compressing the signal in the frequency domain simultaneously. The traditional wavelet denoising for 1D functions suggests that using shrinkage at high-pass coefficients can effectively filter out the Gaussian white noise in the mean square error sense. This is also true for our case when we embed shrinkage in graph convolution. In Figure 3, the framelet convolution with shrinkage activation (UFGCONV-S) outperforms the ReLU case (UFGCONV-R) for reducing node and structure noises. Both methods surpass the classic spatial convolutions GCN (Kipf & Welling, 2017) and GAT (Veličković et al., 2018). Apart from denoising, graph convolution with shrinkage diminishes the proportion of non-zero framelet coefficients while maintains a comparable learning performance. We define the *compression ratio* for a shrinkage framelet convolutional layer as the ratio of the number of non-zero coefficients after and before shrinkage. Tables 1 and 2 show that UFGCONV-S compresses up to 70% non-zero coefficients with top performance for various node classification tasks.

Framelet Spectrum, Training Loss and Network Capacity The coefficients after shrinkage activation are proportional to the framelet power spectrum at the coefficient scale. We thus let the threshold level σ proportionate upon the framelet energy $\|\mathcal{W}_{r,j}X\|^2$ ($r > 0$) for high-passes. For example, the framelet spectrum curves in training for the UFGCONV of Figure 2 show a higher magnitude order of the low-pass (column 3) than those of high-passes (columns 4-5). This is because coefficients in high-passes reflect more detailed characteristics than in low-pass. Compared with the ReLU case (row 1), the shrinkage activation (row 2) filters out some high-pass coefficients in graph convolution, which results in much smaller framelet spectra for high-passes. In contrast, the low-pass shrinkage involves no cutoff, and the energy is less distinguishable from the ReLU case.

The training loss curve of each output feature (column 1) indicates that shrinkage allows for more stable training, with a monotonically decreasing loss. The splitting in low-pass and high-passes for loss suggests a more flexible and precise control of the training. It also opens the possibility of designing a weighted new loss taking account of framelet spectrum. Moreover, the l_2 norm of the network filter θ (column 2) has an increasing trend for the low-pass part and a decreasing trend for the high-pass parts during training. This observation is identical to the spectral bias (Rahaman et al., 2019), whereby the fitting capacity of the framelet convolutional GNN with either ReLU or shrinkage activation

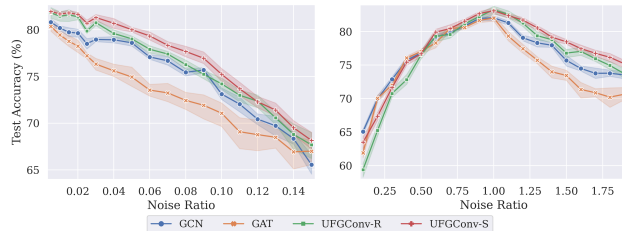


Figure 3. Node (left) and structure (right) perturbation analysis on Cora. Results from the other two datasets are in Appendix. The framelet dilation and scale level are both set to default value 2, and the optimal threshold σ for shrinkage is searched from $\{0.05, 0.1, 0.15\}$. Framelet convolution with shrinkage performs the best under both node and structure perturbations.

comes from the low-pass channel.

6. Robustness of Framelet Convolution under Feature and Structure Noises

Real-world data are usually noisy. For instance, graph data are sometimes polluted due to adversarial attacks as GNNs exchange node information (Xu et al., 2020), where node feature and graph structure could both be perturbed.

Our shrinkage framelet convolution has a motivation from the image deconvolution (or image restoration) model. Given the original and observed (degraded) image features x and y , one defines $y = Hx + \epsilon$, where H represents the convolution matrix of x . The noise ϵ is assumed multivariate Gaussian. In statistical formulation, the model solves

$$\hat{x} = \arg \min_x \{ \log(\Pr(y|x)) - \text{Pen}(x) \}$$

with a penalty function Pen. On the real axis, wavelets are critical in restoring x from noisy y (Figueiredo & Nowak, 2003; Cai et al., 2012; Dong & Shen, 2013; Shen, 2010). With wavelet transform Φ , the maximum penalized likelihood estimator (MPLE) takes the form

$$\hat{x} = \Phi^\top D \Phi y, \quad (4)$$

where D acts as a denoising operation of Φy .

The above wavelet-based convolution can be generalized to graph data restoration with x and y replaced by clean and distorted features on the graph. That is,

$$y = Px + \epsilon, \quad (5)$$

where ϵ is the entry-wise noise. The linear transform P is a permutation of node features or a change of the adjacency matrix. Similar models using graph Fourier spectrum were considered in image restoration, see Cheung et al. (2018, Sect V.B) and Milanfar (2013).

Similar to (4), our graph convolution replaces Φ with the graph framelet transform \mathcal{W} , and D with the trainable network filter θ . Our shrinkage activation is partly motivated by

LASSO (Tibshirani, 1996). The latter uses shrinkage thresholding to denoise the signal in (5) in high dimensions. Based on this connection, our proposed framelet convolution in Section 4 has a good potential against perturbation from (5). We test UFGCONV with perturbed nodes and edges in citation network datasets (Cora, Citeseer and Pubmed). The noise ϵ follows Bernoulli distribution to match binary node features of the tasks. That is, we randomly change node feature or edge weight from 0/1 to its opposite 1/0. Figure 3 reports the test accuracy of UFGCONV, GCN and GAT. The x -axis indicates the proportion of the distorted nodes or edges (which is equivalent to the signal-to-noise ratio (SNR)). As the SNR increases, UFGCONV behaves well ahead of GCN and GAT while with higher test accuracy and smaller variance. The slightly lower performance of UFGCONV only occurs when more than 50% edges are removed, where misinformation dominates the graph structure. Moreover, the shrinkage thresholding in all situations manages to filter out more noise and thus achieves higher performance. This example illustrates the effectiveness of framelet convolution in predicting node property with node features or structure of graphs that are distorted.

7. Framelet Pooling

Graph pooling is a critical ingredient of GNNs when the model is predicting graph-level properties with a constant feature dimension but varying graph size and structure. We propose *framelet pooling* for GNNs using framelet transforms. As an illustrative example, we use 2 scale levels for framelet decomposition. Similar to the graph convolution in Section 4, given a graph with feature matrix $X \in \mathbb{R}^{N \times d}$, we can obtain a set of framelet coefficients $\mathcal{W}_{r,j}f$ including one low pass $\mathcal{W}_{0,2}f$ at level 2 and two high passes $\mathcal{W}_{1,1}f$ and $\mathcal{W}_{1,2}f$ at levels 1 and 2, respectively. Each scale-wise framelet coefficient is an $N \times d$ real-valued matrix, and its i th feature column $(\mathcal{W}_{r,j}X)_i$ for $i = 1, \dots, d$ would be aggregated by the sum, or the sum of squares of the elements. The two aggregation methods correspond to two framelet pooling strategies (see below). The calculation compresses the $N \times d$ coefficients to a d -dimensional vector, and the pooled output from the three framelet coefficients results in 3 d -dimensional vectors. Figure 4 visualizes the computational flowchart for our pooling model.

The framelet pooling benefits the network training by employing the information from multi-scales, as all scales in the framelet representation of the graph signal is taken into account. Depending on how we aggregate the framelet coefficients, we distinguish the two strategies by UFGPOOL-SUM or UFGPOOL-SPECTRUM. The latter aggregates the nodes by the wavelet (power) spectrum (that is, the sum of absolute squares of framelet coefficients over nodes, $\sum_{p \in V} |v_{j,p}|^2$ and $\sum_{p \in V} |w_{j,p}^r|^2$). In this way, the total in-

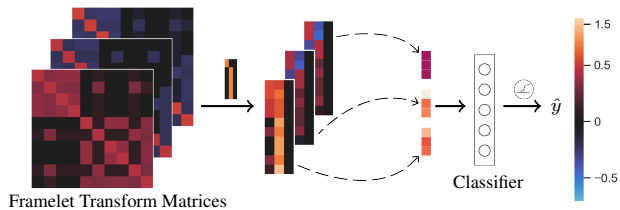


Figure 4. Framelet pooling for graph property prediction. The three framelet transform matrices are retrieved from Figure 1 with the same protein sample and parameter setting. The scale-wise framelet coefficients are aggregated to three vectors by sum or sum of squares (framelet spectrum). The (1 low pass and 2 high passes) vectors are then concatenated as the readout for the classifier.

formation of the graph signal X^{in} is well-conserved after the pooling. The sum of wavelet power spectrum is equal to the total energy of the signal, that is, $\|X^{pooled}\| = \|X^{in}\|$ (see Theorem 1(iii) in Appendix). We present empirical evidence of the precedence of the proposed framelet pooling over existing graph pooling methods in Section 8.2.

8. Experiments

In this section, we show a variety of numerical tests for our framelet convolution and pooling. Section 8.1 tests the performance of framelet convolution (UFGCONV) on node classification benchmarks. Section 8.2 studies the ablation for the proposed framelet pooling (UFGPOOL). Section 8.3 provides a sensitivity analysis for UFGCONV in terms of dilation and scale. All experiments run in PyTorch on NVIDIA® Tesla V100 GPU with 5,120 CUDA cores and 16GB HBM2 mounted on an HPC cluster.

8.1. Framelet Convolution for Node Classification

We test UFGCONV with both ReLU and shrinkage activations on four node classification datasets. We denote the two variants by UFGCONV-R and UFGCONV-S.

Dataset The first experiment of node classification tasks is conducted on **Cora**, **Citeseer** and **Pubmed**, which are three benchmark citation networks. Moreover, we employ **ogbn-arxiv** from open graph benchmark **OGB** (Hu et al., 2020) to illustrate the power of our framelet convolution on large-scale graph-structured data.

Setup We design our UFGCONV model with two convolution layers for learning the graph embedding, the output of which is proceeded by a softmax activation for final prediction. Most hyperparameters are set to default, except for learning rate, weight decay, hidden units and dropout ratio in training. A grid search is conducted for fine tuning on these hyperparameters from the search space detailed in Appendix. Both methods are trained with the ADAM optimizer. The maximum number of epochs is 200 for citation

Table 1. Test accuracy (in percentage) for citation networks with standard deviation after \pm . Compression ratio in (green) is the ratio of numbers of nonzero coefficients after and before shrinkage, and is with threshold level $\sigma = 1$.

Method	Cora	Citeseer	Pubmed
MLP	55.1	46.5	71.4
DEEPWALK	67.2	43.2	65.3
SPECTRAL	73.3	58.9	73.9
CHEBYSHEV	81.2	69.8	74.4
GCN	81.5	70.3	79.0
GWNN	82.8	71.7	79.1
GAT	83.0 \pm 0.7	72.5 \pm 0.7	79.0 \pm 0.3
MPNN	78.0 \pm 1.1	64.0 \pm 1.9	75.6 \pm 1.0
GRAPHSAGE	74.5 \pm 0.8	67.2 \pm 1.0	76.8 \pm 0.6
LANCZOSNET	79.5 \pm 1.8	66.2 \pm 1.9	78.3 \pm 0.3
DCNN	79.7 \pm 0.8	69.4 \pm 1.3	76.8 \pm 0.8
UFGCONV-S	83.0 \pm 0.5	71.0 \pm 0.6	79.4 \pm 0.4
(Compression)	(47.7)	(39.0)	(27.7)
UFGCONV-R	83.6 \pm 0.6	72.7 \pm 0.6	79.9 \pm 0.1

† The top three are highlighted by **First, Second, Third**.

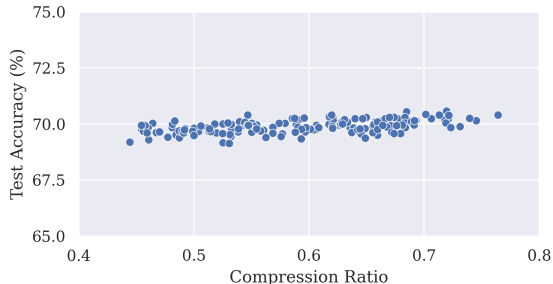


Figure 5. Trade-off between compression ratio and test accuracy in UFGCONV-S with ogbn-arxiv.

networks and 500 for ogbn-arxiv. All the datasets follow the standard public split and processing rules. The average test accuracy and its standard deviation come from 10 runs.

Baseline The UFGCONV-R and UFGCONV-S are compared against other methods for node classification tasks. We consider multiple baseline models that are applicable to the tasks. For citation networks, the reported accuracy are retrieved from public results: MLP, DEEPWALK (Perozzi et al., 2014), CHEBYSHEV (Defferrard et al., 2016a) and GCN (Kipf & Welling, 2017) from Kipf & Welling (2017); SPECTRAL CNN (Bruna et al., 2014) and GWNN (Xu et al., 2019a); MPNN (Gilmer et al., 2017), GRAPH-SAGE (Hamilton et al., 2017), LANCZOSNET (Liao et al., 2019) and DCNN (Singer et al., 2009) from Liao et al. (2019); and GAT (Veličković et al., 2018) from their authors. For ogbn-arxiv, we also compared with NODE2VEC (Grover & Leskovec, 2016), GRAPHZOOM (Deng et al., 2020), P&L+C&S (Huang et al., 2021), DEEPERGCN (Li et al., 2020a), SIGN (Rossi et al., 2020) and GAAN (Zhang et al., 2018a) from the OGB leaderboard.

Table 2. Test accuracy (in percentage) for ogbn-arxiv with standard deviation after \pm . The compression ratio for UFGCONV-S with shrinkage threshold level $\sigma = 1$ is **64.2%**.

Method	Test Acc.	Val. Acc.	#Params
MLP	55.50 \pm 0.23	57.65 \pm 0.12	110, 120
NODE2VEC	70.07 \pm 0.13	71.29 \pm 0.13	21, 818, 792
GRAPHZOOM	71.18 \pm 0.18	72.20 \pm 0.07	8, 963, 624
P&L + C&S	71.26 \pm 0.01	73.00 \pm 0.01	5, 160
GRAPHSAGE	71.49 \pm 0.27	72.77 \pm 0.17	218, 664
GCN	71.74 \pm 0.29	73.00 \pm 0.17	142, 888
DEEPERGCN	71.92 \pm 0.17	72.62 \pm 0.14	491, 176
SIGN	71.95 \pm 0.11	73.23 \pm 0.06	3, 566, 128
GAAN	71.97 \pm 0.18	–	1, 471, 506
UFGCONV-S	70.04 \pm 0.22	71.04 \pm 0.11	1, 633, 183
UFGCONV-R	71.97 \pm 0.12	73.21 \pm 0.05	1, 633, 183

† The top three are highlighted by **First, Second, Third**.

Results We report the accuracy score in percentage with the top-3 highlighted in Tables 1 and 2. For UFGCONV-S, we also report the compression ratio for shrinkage (in green). In Table 1, the UFGCONV-R method achieves the highest prediction accuracy among all baseline models. The learned UFGCONV-S with threshold level $\sigma = 1$ trims up to 50% information but still obtains the top-3 rank on two tasks. A similar outstanding performance is reported in Table 2 for the ogbn-arxiv dataset, where the UFGCONV-R again ranks first with a moderate number of parameters, and the UFGCONV-S with threshold $\sigma = 1$ achieves a comparable accuracy at 70% using 64.2% information.

We test UFGCONV-S with different compression ratios. Ideally, a high test accuracy is preferable to pair with a low compression ratio, and the change in accuracy should be minimal due to the insensitivity of our model to the hyper-parameters. However, as shown in Figure 5, an increasing compression ratio generally results in a slightly higher prediction accuracy as more coefficients for the framelet representation is used by the convolution.

8.2. Framelet Pooling for Graph Property Prediction

The second experiment evaluates two framelet pooling methods, UFGPOOL-SUM and UFGPOOL-SPECTRUM, by ablation studies on graph classification and regression tasks.

Dataset We select six benchmarks to test the proposed pooling strategies, including four graph classification tasks with moderate sample sizes, one regression task, and one large-scale classification task. First five tasks use **TU-Dataset benchmarks** (Morris et al., 2020), including **D&D** (Dobson & Doig, 2003; Shervashidze et al., 2011), **PROTEINS** (Dobson & Doig, 2003; Borgwardt et al., 2005) to categorize proteins into enzyme and non-enzyme structures; **NC11** (Wale et al., 2008) to identify chemical compounds that block lung cancer cells; **Mutagenicity** (Kazius et al.,

Table 3. Performance comparison for graph property prediction. **QM7** is a regression task in MSE; **ogbg-molhiv** is a classification task in ROC-AUC in percentage; others are for classification in test accuracy in percentage. The value after \pm is standard deviation.

Datasets	PROTEINS	Mutagenicity	D&D	NCI1	ogbg-molhiv	QM7
TOPKPOOL	73.48 \pm 3.57	79.84 \pm 2.46	74.87 \pm 4.12	75.11 \pm 3.45	78.14 \pm 0.62	175.41 \pm 3.16
ATTENTION	73.93 \pm 5.37	80.25 \pm 2.22	77.48 \pm 2.65	74.04 \pm 1.27	74.44 \pm 2.12	177.99 \pm 2.22
SAGPOOL	75.89 \pm 2.91	79.86 \pm 2.36	74.96 \pm 3.60	76.30 \pm 1.53	75.26 \pm 2.29	41.93 \pm 1.14
SUM	74.91 \pm 4.08	80.69 \pm 3.26	78.91 \pm 3.37	76.96 \pm 1.70	77.41 \pm 1.16	42.09 \pm 0.91
MAX	73.57 \pm 3.94	78.83 \pm 1.70	75.80 \pm 4.11	75.96 \pm 1.82	78.16 \pm 1.33	177.48 \pm 4.70
MEAN	73.13 \pm 3.18	80.37 \pm 2.44	76.89 \pm 2.23	73.70 \pm 2.55	78.21 \pm 0.90	177.49 \pm 4.69
UFGPOOL-SUM	77.77 \pm 2.60	81.59 \pm 1.40	80.92 \pm 1.68	77.88 \pm 1.24	78.80 \pm 0.56	41.74 \pm 0.84
UFGPOOL-SPECTRUM	77.23 \pm 2.40	82.05 \pm 1.28	79.83 \pm 1.88	77.54 \pm 2.24	78.36 \pm 0.77	41.67 \pm 0.95

† The top three are highlighted by **First**, **Second**, **Third**.

2005; Riesen & Bunke, 2008) to recognize mutagenic molecular compounds for potentially marketable drug; and **QM7** (Blum & Reymond, 2009; Rupp et al., 2012) to predict atomization energy value of molecules. The dataset **ogbg-molhiv** (Hu et al., 2020) is for large-scale molecule classification.

Setup The network architecture for all baseline models is fixed to two convolutional layers followed by one pooling layer. The graph convolution for the five **TU**Datasets uses the GCN model, and for **ogbg-molhiv** uses GIN with virtual nodes (Ishiguro et al., 2019). Given graph representations, the prediction is made by a two-layer MLP, where the hidden unit is identical to that of the convolutional layer. The hyperparameters (learning rate, weight decay, number of hidden units in each convolutional layer, and dropout ratio in the readout layer) are fine-tuned with grid search.

Each dataset is split into training, validation and test sets by 80%, 10% and 10%. The training stops when the validation loss stops improving for 20 consecutive epochs or reaching maximum 200 epochs. All results are averaged over 10 repetitions. The classification tasks report mean test accuracy for **TU**Dataset and ROC-AUC score for **ogbg-molhiv**. The regression task on **QM7** reports mean square error (MSE).

Baseline We compare our framelet pooling (UFGPOOL-SUM and UFGPOOL-SPECTRUM) with six baseline methods that are capable for global pooling to verify the effectiveness of the learned graph representation. The baselines include TOPKPOOL (Gao & Ji, 2019; Cangea et al., 2018), ATTENTIONPOOL (Li et al., 2016), SAGPOOL (Lee et al., 2019), as well as the classic SUM, MEAN and MAX pooling.

Results Table 3 summarizes the performance comparison. Our UFGPOOL methods outperform other methods on all datasets. Specifically, UFGPOOL-SUM achieves the top accuracy in four out of six datasets, and the second best accuracy in the other two, where the top performance is achieved by UFGPOOL-SPECTRUM. We also observe that UFGPOOL-SPECTRUM performs better on small molecules prediction: **Mutagenicity**, **QM7** and **ogbg-molhiv**. This

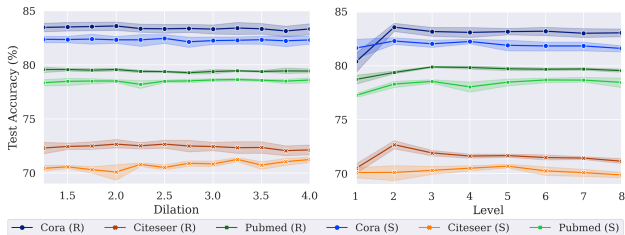


Figure 6. Sensitivity analysis for dilation (left) and scale level (right) with UFGCONV-R and UFGCONV-S on citation networks.

precedence might come from encoding the multi-scale signal energy to the network where the framelet spectra capture the practically significant features of molecular data.

8.3. Sensitivity Analysis

This section analyses the sensitivity of UFGCONV-R and UFGCONV-S on the hyperparameters dilation and scale level in the framelet system. The experiment is conducted on **Cora**, **Citeseer** and **Pubmed**. The dilation analysis selects values from 1.25 to 4 with step 0.25, and the scale levels in those three datasets are fixed to 2, 2 and 3 respectively. For the scale level analysis, the values are set from 1 to 8 with step 1, and the dilation stays at default 2. All other hyperparameters are tuned in the same way as Section 8.1. We use shrinkage threshold $\sigma = 1$ for UFGCONV-S.

From Figure 6, we can observe that changing dilation or scale level does not drastically impact on the accuracy for either method. In particular, the mean test accuracy is stable over all dilation values and reaches the peak with a small scale level (2 for **Cora** & **Citeseer**; 3 for **Pubmed**). For scale level 1, the decreased accuracy is due to the insufficiency of scale and then not salient multiresolution. Thus, we can use dilation 2 and scale level 2 in practice, in which the GNN uses multi-scale framelet analysis to achieve supreme performance with a low computational cost.

9. Conclusion

We explore the adaptation of graph framelets for graph neural networks in this paper. As a multi-scale graph representa-

tion method, framelet transforms link graph neural networks and signal processing. In many node-level or graph-level tasks, framelet convolutions can reduce both feature and structure noises. We also introduce shrinkage activation that thresholds high-pass coefficients in framelet convolution, which strengthens the network denoising capability and simultaneously compresses graph signal at a remarkable rate. Moreover, we design graph pooling using framelet spectra at low and high passes. The proposed framelet convolutions with both ReLU and shrinkage surpass typical spatial-based and spectral-based graph convolutions on most benchmarks, and the framelet pooling outperforms baselines on a variety of graph property prediction tasks.

Acknowledgements

YW and GM have been supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant n° 757983). ML acknowledges supports from the National Natural Science Foundation of China (No. 61802132) and the “Qianjiang Talent Program D” of Zhejiang Province. This project was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. This research was also undertaken with the assistance of resources and services from the GPU cluster Cobra of Max Planck Society. The authors would like to acknowledge support from the UNSW Resource Allocation Scheme managed by Research Technology Services at UNSW Sydney. The authors also acknowledge the technical assistance provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.

References

- Balcilar, M., Renton, G., Héroux, P., Gaüzère, B., Adam, S., and Honeine, P. Analyzing the expressive power of graph neural networks in a spectral perspective. In *ICLR*, 2021.
- Behjat, H., Richter, U., Van De Ville, D., and Sörnmo, L. Signal-adapted tight frames on graphs. *IEEE Transactions on Signal Processing*, 64(22):6017–6029, 2016.
- Blum, L. C. and Raymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *Journal of the American Chemical Society*, 131:8732, 2009.
- Borgwardt, K. M., Ong, C. S., Schönauer, S., Vishwanathan, S. V. N., Smola, A. J., and Kriegel, H.-P. Protein function prediction via graph kernels. *Bioinformatics*, 21(suppl_1):i47–i56, 2005.
- Brauchart, J., Dick, J., Saff, E., Sloan, I., Wang, Y., and Womersley, R. Covering of spheres by spherical caps and worst-case error for equal weight cubature in sobolev spaces. *Journal of Mathematical Analysis and Applications*, 431(2):782 – 811, 2015. ISSN 0022-247X.
- Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y. Spectral networks and locally connected networks on graphs. In *ICLR*, 2014.
- Cai, J.-F., Dong, B., Osher, S., and Shen, Z. Image restoration: total variation, wavelet frames, and beyond. *Journal of the American Mathematical Society*, 25(4):1033–1089, 2012.
- Cangea, C., Veličković, P., Jovanović, N., Kipf, T., and Liò, P. Towards sparse hierarchical graph classifiers. In *NeurIPS Workshop on Relational Representation Learning*, 2018.
- Chen, M., Wei, Z., Huang, Z., Ding, B., and Li, Y. Simple and deep graph convolutional networks. In *ICML*, pp. 1725–1735. PMLR, 2020.
- Cheung, G., Magli, E., Tanaka, Y., and Ng, M. K. Graph spectral image processing. *Proceedings of the IEEE*, 106(5):907–930, 2018.
- Chui, C. K., Filbir, F., and Mhaskar, H. N. Representation of functions on big data: graphs and trees. *Applied and Computational Harmonic Analysis*, 38(3):489–509, 2015.
- Cohen, A., Daubechies, I., Jawerth, B., and Vial, P. Multiresolution analysis, wavelets and fast algorithms on an interval. *Comptes rendus de l’Académie des sciences. Série I, Mathématique*, 316(5):417–421, 1993.
- Crovella, M. and Kolaczyk, E. Graph wavelets for spatial traffic analysis. In *IEEE INFOCOM 2003*, volume 3, pp. 1848–1857. IEEE, 2003.
- Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In *NIPS*, pp. 3844–3852, 2016a.
- Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In *NIPS*, pp. 3844–3852, 2016b.
- Deng, C., Zhao, Z., Wang, Y., Zhang, Z., and Feng, Z. Graphzoom: A multi-level spectral approach for accurate and scalable graph embedding. In *ICLR*, 2020.
- Dobson, P. D. and Doig, A. J. Distinguishing enzyme structures from non-enzymes without alignments. *Journal of Molecular Biology*, 330(4):771–783, 2003.
- Dong, B. Sparse representation on graphs by tight wavelet frames and applications. *Applied and Computational Harmonic Analysis*, 42(3):452–479, 2017.

- Dong, B. and Shen, Z. *MRA-Based Wavelet Frames and Applications*. IAS Lecture Note Series, 06 2013.
- Donoho, D. L. Wavelet shrinkage and wvd: a 10-minute tour. In *Presented on the International Conference on Wavelets and Applications, Toulouse, France*, 1992.
- Donoho, D. L. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.
- Donoho, D. L. and Johnstone, J. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.
- Donoho, D. L., Johnstone, I. M., Kerkycharian, G., and Picard, D. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(2):301–337, 1995.
- Efron, B. and Morris, C. Data analysis using Stein’s estimator and its generalizations. *Journal of the American Statistical Association*, 70(350):311–319, 1975.
- Figueiredo, M. A. and Nowak, R. D. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003.
- Gao, H. and Ji, S. Graph U-nets. In *ICML*, 2019.
- Gavish, M., Nadler, B., and Coifman, R. R. Multiscale wavelets on trees, graphs and high dimensional data: theory and applications to semi supervised learning. In *ICML*, pp. 367–374, 2010.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *ICML*, pp. 1263–1272, 2017.
- Glorot, X. and Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*, pp. 249–256, 2010.
- Grover, A. and Leskovec, J. node2vec: Scalable feature learning for networks. In *ACM SIGKDD*, pp. 855–864, 2016.
- Hamilton, W., Ying, Z., and Leskovec, J. Inductive representation learning on large graphs. In *NIPS*, pp. 1024–1034, 2017.
- Hammond, D. K., Vandergheynst, P., and Gribonval, R. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.
- Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., and Leskovec, J. Open graph benchmark: Datasets for machine learning on graphs. In *NeurIPS*, 2020.
- Huang, Q., He, H., Singh, A., Lim, S.-N., and Benson, A. R. Combining label propagation and simple models out-performs graph neural networks. *ICLR*, 2021.
- Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, pp. 448–456, 2015.
- Ishiguro, K., Maeda, S.-i., and Koyama, M. Graph warp module: an auxiliary module for boosting the power of graph neural networks. *arXiv:1902.01020*, 2019.
- ISO. ISO/IEC 15444-1:2019 Information technology - JPEG 2000 image coding system - Part 1: Core coding system. 2019.
- Kazius, J., McGuire, R., and Bursi, R. Derivation and validation of toxicophores for mutagenicity prediction. *Journal of Medicinal Chemistry*, 48(1):312–320, 2005.
- Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.
- Knyazev, B., Taylor, G. W., and Amer, M. R. Understanding attention and generalization in graph neural networks. In *NeurIPS*, 2019.
- Lee, J., Lee, I., and Kang, J. Self-attention graph pooling. In *ICML*, 2019.
- Li, G., Xiong, C., Thabet, A., and Ghanem, B. DeeperGCN: All you need to train deeper GCNs. *arXiv:2006.07739*, 2020a.
- Li, M., Ma, Z., Wang, Y. G., and Zhuang, X. Fast haar transforms for graph neural networks. *Neural Networks*, pp. 188–198, 2020b.
- Li, Y., Tarlow, D., Brockschmidt, M., and Zemel, R. Gated graph sequence neural networks. In *ICLR*, 2016.
- Liao, R., Zhao, Z., Urtasun, R., and Zemel, R. Lanczosnet: Multi-scale deep graph convolutional networks. In *ICLR*, 2019.
- Ma, Z., Xuan, J., Wang, Y. G., Li, M., and Liò, P. Path integral based convolution and pooling for graph neural networks. In *NeurIPS*, volume 33, pp. 16433–16445, 2020.
- Maggioni, M. and Mhaskar, H. H. Diffusion polynomial frames on metric measure spaces. *Applied and Computational Harmonic Analysis*, 24(3):329 – 353, 2008.
- Mallat, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

- Milanfar, P. A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE Signal Processing Magazine*, 30(1):106–128, 2013.
- Morris, C., Kriege, N. M., Bause, F., Kersting, K., Mutzel, P., and Neumann, M. TUDataset: A collection of benchmark datasets for learning with graphs. In *ICML Workshop “Graph Representation Learning and Beyond”*, 2020.
- Perozzi, B., Al-Rfou, R., and Skiena, S. DeepWalk: Online learning of social representations. In *KDD*, pp. 701–710, 2014.
- Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., and Courville, A. On the spectral bias of neural networks. In *ICML*, pp. 5301–5310. PMLR, 2019.
- Riesen, K. and Bunke, H. Iam graph database repository for graph based pattern recognition and machine learning. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pp. 287–297. Springer, 2008.
- Rossi, E., Frasca, F., Chamberlain, B., Eynard, D., Bronstein, M., and Monti, F. SIGN: Scalable inception graph neural networks. In *ICML Workshop “Graph Representation Learning and Beyond”*, 2020.
- Rupp, M., Tkatchenko, A., Müller, K.-R., and von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108:058301, 2012.
- Shen, Z. Wavelet frames and image restorations. In *the International Congress of Mathematicians (ICM)*, pp. 2834–2863, 2010.
- Shervashidze, N., Schweitzer, P., Van Leeuwen, E. J., Mehlhorn, K., and Borgwardt, K. M. Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research*, 12(77):2539–2561, 2011.
- Singer, A., Shkolnisky, Y., and Nadler, B. Diffusion interpretation of nonlocal neighborhood filters for signal denoising. *SIAM Journal on Imaging Sciences*, 2(1):118–139, 2009.
- Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. Graph attention networks. In *ICLR*, 2018.
- Vignac, C., Loukas, A., and Frossard, P. Building powerful and equivariant graph neural networks with structural message-passing. In *NeurIPS*, volume 33, pp. 14154–14166, 2020.
- Wale, N., Watson, I. A., and Karypis, G. Comparison of descriptor spaces for chemical compound retrieval and classification. *Knowledge and Information Systems*, 14(3):347–375, 2008.
- Wang, X., Zhu, M., Bo, D., Cui, P., Shi, C., and Pei, J. Am-gcn: Adaptive multi-channel graph convolutional networks. In *ACM SIGKDD*, pp. 1243–1253, 2020a.
- Wang, Y. G. and Zhuang, X. Tight framelets and fast framelet filter bank transforms on manifolds. *Applied and Computational Harmonic Analysis*, 2018.
- Wang, Y. G. and Zhuang, X. Tight framelets on graphs for multiscale data analysis. In *Wavelets and Sparsity XVIII*, volume 11138, pp. 100 – 111. SPIE, 2019.
- Wang, Y. G., Li, M., Ma, Z., Montufar, G., Zhuang, X., and Fan, Y. Haar graph pooling. In *ICML*, pp. 9952–9962. PMLR, 2020b.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Yu, P. S. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021.
- Xu, B., Shen, H., Cao, Q., Qiu, Y., and Cheng, X. Graph wavelet neural network. In *ICLR*, 2019a.
- Xu, H., Ma, Y., Liu, H.-C., Deb, D., Liu, H., Tang, J.-L., and Jain, A. K. Adversarial attacks and defenses in images, graphs and text: A review. *International Journal of Automation and Computing*, 17(2):151–178, 2020.
- Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? In *ICLR*, 2019b.
- Zhang, J., Shi, X., Xie, J., Ma, H., King, I., and Yeung, D.-Y. GaAN: Gated attention networks for learning on large and spatiotemporal graphs. In *UAI*, pp. 339–349, 2018a.
- Zhang, M., Cui, Z., Neumann, M., and Chen, Y. An end-to-end deep learning architecture for graph classification. In *AAAI*, 2018b.
- Zheng, X., Zhou, B., Li, M., Wang, Y. G., and Gao, J. MathNet: Haar-like wavelet multiresolution-analysis for graph representation and learning. *arXiv:2007.11202*, 2020a.
- Zheng, X., Zhou, B., Wang, Y. G., and Zhuang, X. Decimated framelet system on graphs and fast G-framelet transforms. *arXiv:2012.06922*, 2020b.