Graph Convolution Networks

SeungWoo Yu

Seoul National University

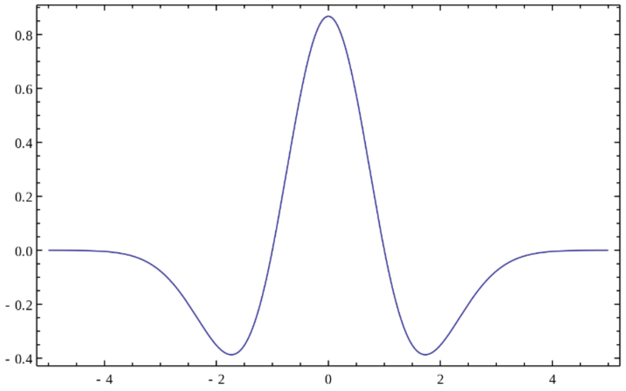
Wavelet Transform

$$\mathcal{W}[f](a,b) = <f, \psi^{a,b}> = \int f(x)\overline{\psi^{a,b}}dx$$

$$\psi^{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \qquad \|\psi^{a,b}\| = \|\psi\|$$

$$\psi_{m,n}(x) = \frac{1}{\sqrt{a_0^m}}\psi\left(\frac{x - nb_0 a_0^m}{a_0^m}\right)$$

$$\mathcal{W}_f(s,n) = <f, \psi_{s,n}>$$



Mexican hat Wavelet

Graph Wavelet Transform

- Scaling & Translation(Hammond et al,2009)

$$< f, \psi_{s,n} >= \sum_i \sum_l g(s\lambda_l)\, \mathbf{u}_l(n)\, \mathbf{u}_l(i) f(i)$$

- Kernel

$$< f, \psi_{s,n} >= \sum_l g(s\lambda_l)\, \mathbf{u}_l(n) \hat{f}(\lambda_l)$$

- Polynomial Approximation

$$\tilde{W}_f(t,n) = (p(\mathcal{L})f)_n$$

$$\hat{g}(\lambda_i) \cdot \boldsymbol{u}_i^T \boldsymbol{f}$$

Spectral GCN

Graph Wavelet Transform

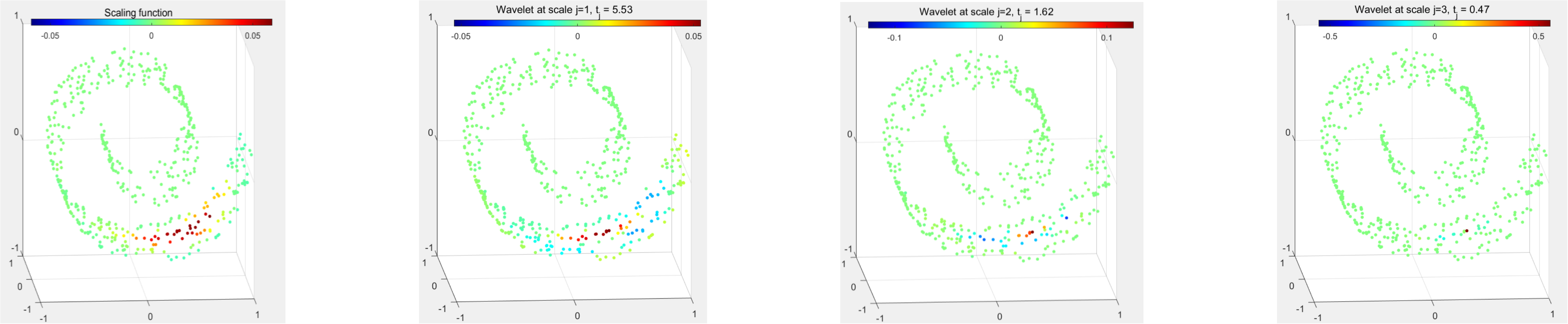
- Polynomial Approximation for fast SGWT(Spectral Graph Wavelet Transform)

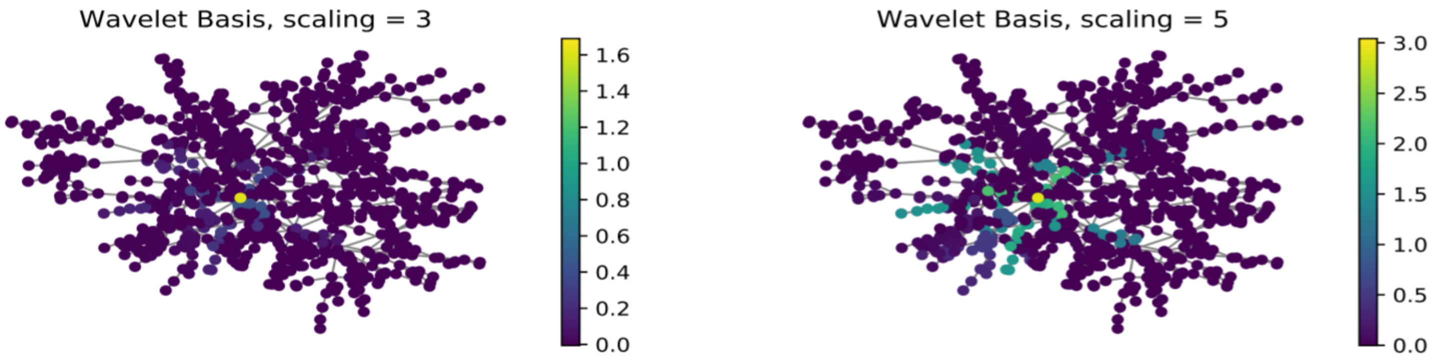
$$\tilde{W}_f(t_j, n) = \left(\frac{1}{2} c_{j,0} f + \sum_{k=1}^{M_j} c_{j,k} \overline{T}_k(\mathcal{L}) f \right)_n$$

$$c_{n,k} = \frac{2}{\pi} \int_0^{\pi} \cos(k\theta) g(t_n(a(\cos(\theta) + 1))) d\theta.$$

Graph Wavelet Neural Network

- Sparseness of wavelet





(a) (b)

Figure 1: Wavelets on an example graph at (a) small scale and (b) large scale.

Experimental Results

- Parameter complexity
  - Two level learning in a layer

$$\text{feature transformation}: \ \boldsymbol{X}^{m'} = \boldsymbol{X}^m \boldsymbol{W},$$

$$\text{graph convolution}: \ \boldsymbol{X}^{m+1} = h(\psi_s \boldsymbol{F}^m \psi_s^{-1} \boldsymbol{X}^{m'}).$$

  - Big O of parameter complexity

$$O(K \times p \times q) \Rightarrow O(K + p \times q)$$

Table 2: Results of Detaching Feature Transformation from Convolution

	Method	Cora	Citeseer	Pubmed
Prediction Accuracy	ChebyNet	81.2%	**69.8%**	74.4%
	Detaching-ChebyNet	**81.6%**	68.5%	**78.6%**
Number of Parameters	ChebyNet	46,080 (K=2)	178,032 (K=3)	24,144 (K=3)
	Detaching-ChebyNet	23,048 (K=4)	59,348 (K=2)	8,054 (K=3)

Experimental Results

- Accuracy

Table 3: Results of Node Classification

Method	Cora	Citeseer	Pubmed
MLP	55.1%	46.5%	71.4%
ManiReg	59.5%	60.1%	70.7%
SemiEmb	59.0%	59.6%	71.7%
LP	68.0%	45.3%	63.0%
DeepWalk	67.2%	43.2%	65.3%
ICA	75.1%	69.1%	73.9%
Planetoid	75.7%	64.7%	77.2%
Spectral CNN	73.3%	58.9%	73.9%
ChebyNet	81.2%	69.8%	74.4%
GCN	81.5%	70.3%	79.0%
MoNet	81.7±0.5%	—	78.8±0.3%
GWNN	**82.8%**	**71.7%**	**79.1%**

- Matrix Density

Table 4: Statistics of wavelet transform and Fourier transform on Cora

	Statistical Property	wavelet transform	Fourier transform
Transform Matrix	Density	2.8%	99.1%
	Number of Non-zero Elements	205,774	7,274,383
Projected Signal	Density	10.9%	100%
	Number of Non-zero Elements	297	2,708

Conclusion

- Graph wavelets are local and sparse
- Graph wavelet transform is computationally efficient
- Convolution is localized in vertex domain
- Detached the feature transformation from convolution
  - Reduce the number of parameters

- My opinion
  - Performance reduction when Scaling factor number is increased
  - Using only two layer