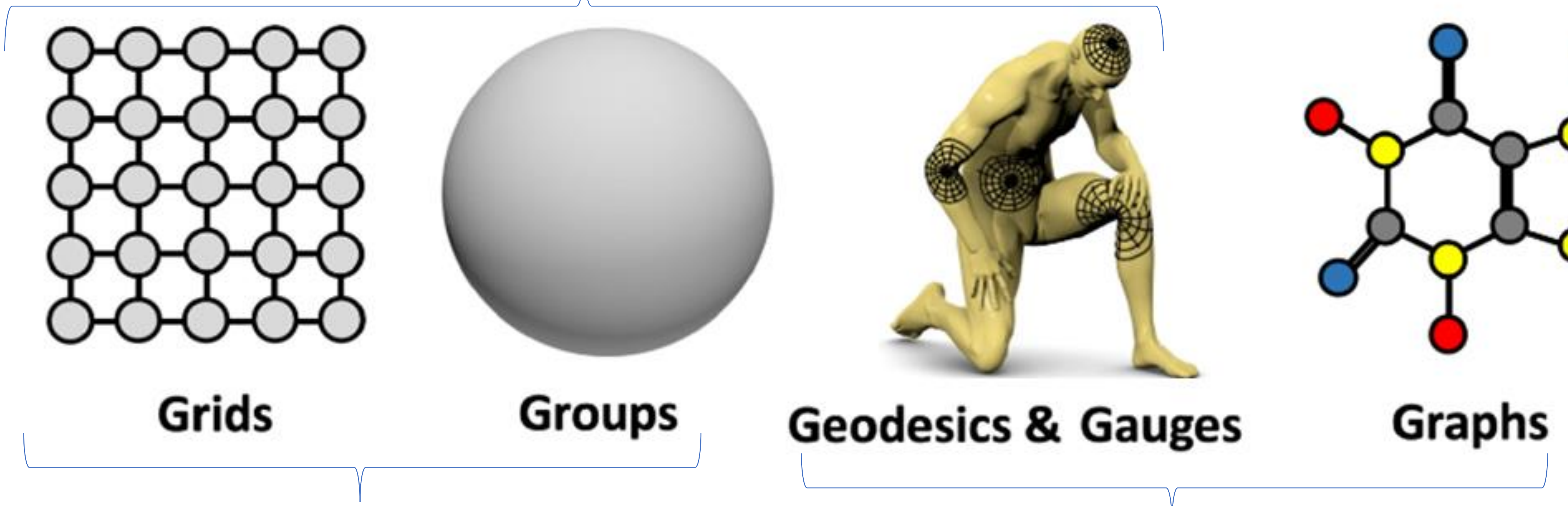
Structural and Spectral Awareness in GNN

Muhammet Balcilar PhD

R&I Researcher at Interdigital, France

Nodes have positional information

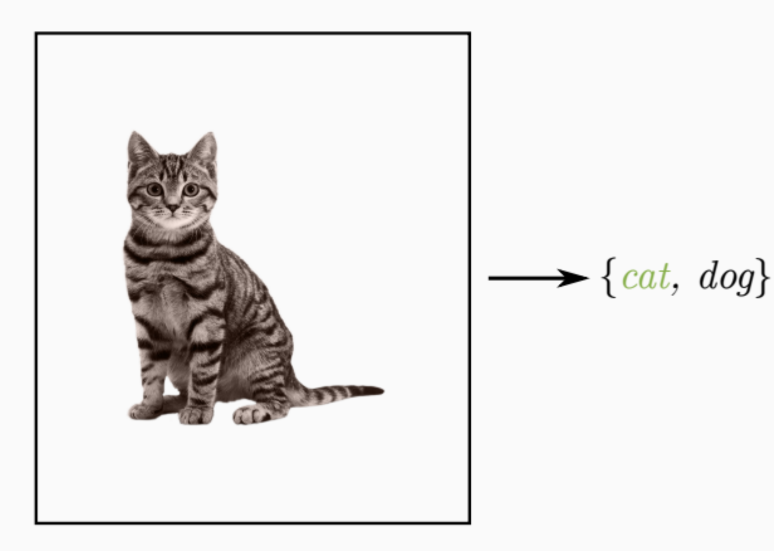


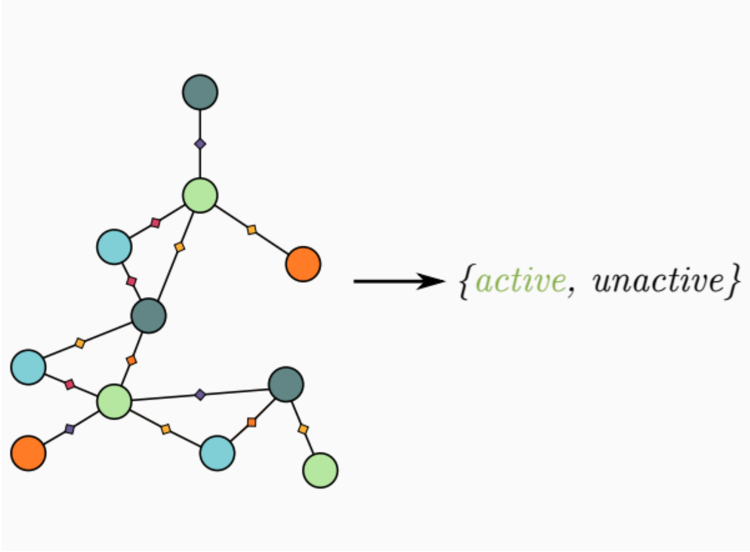
Regular structure

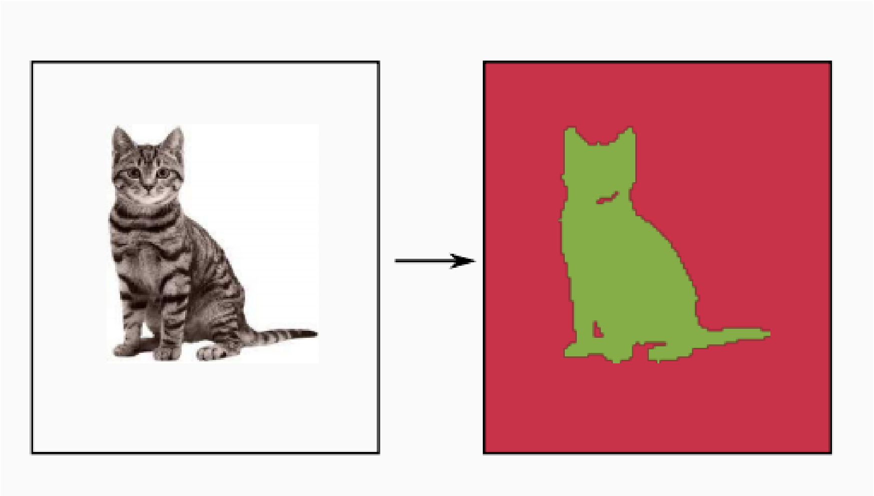
Irregular structure

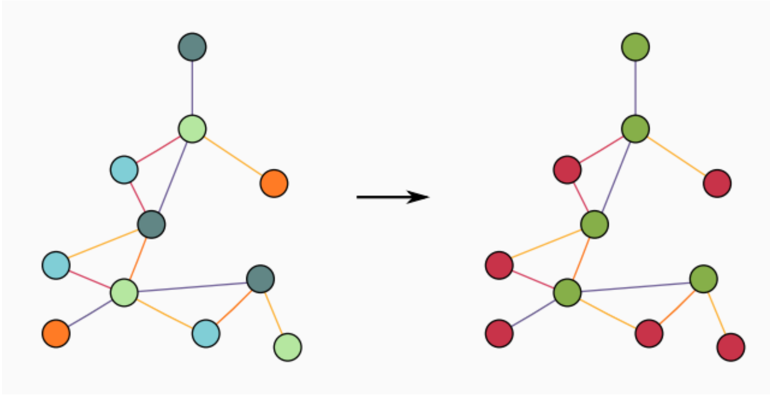
GEOMETRIC DEEP LEARNING

CNN vs GNN

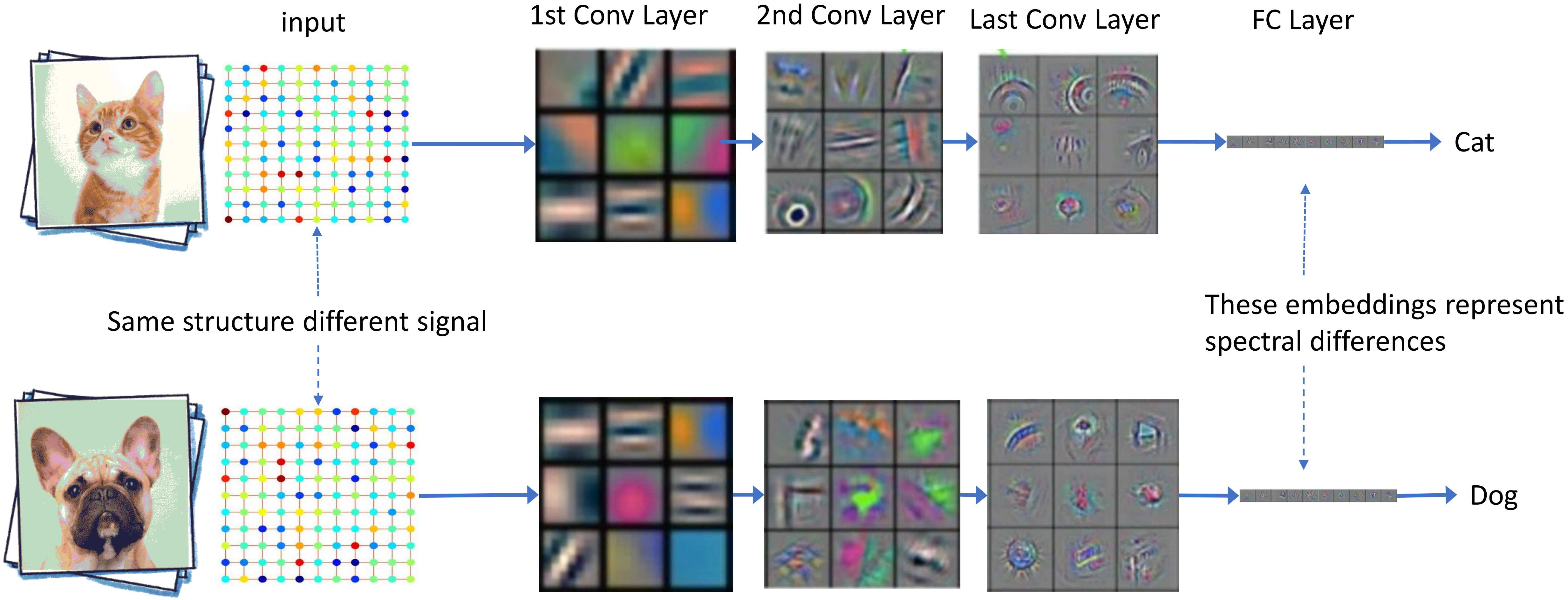






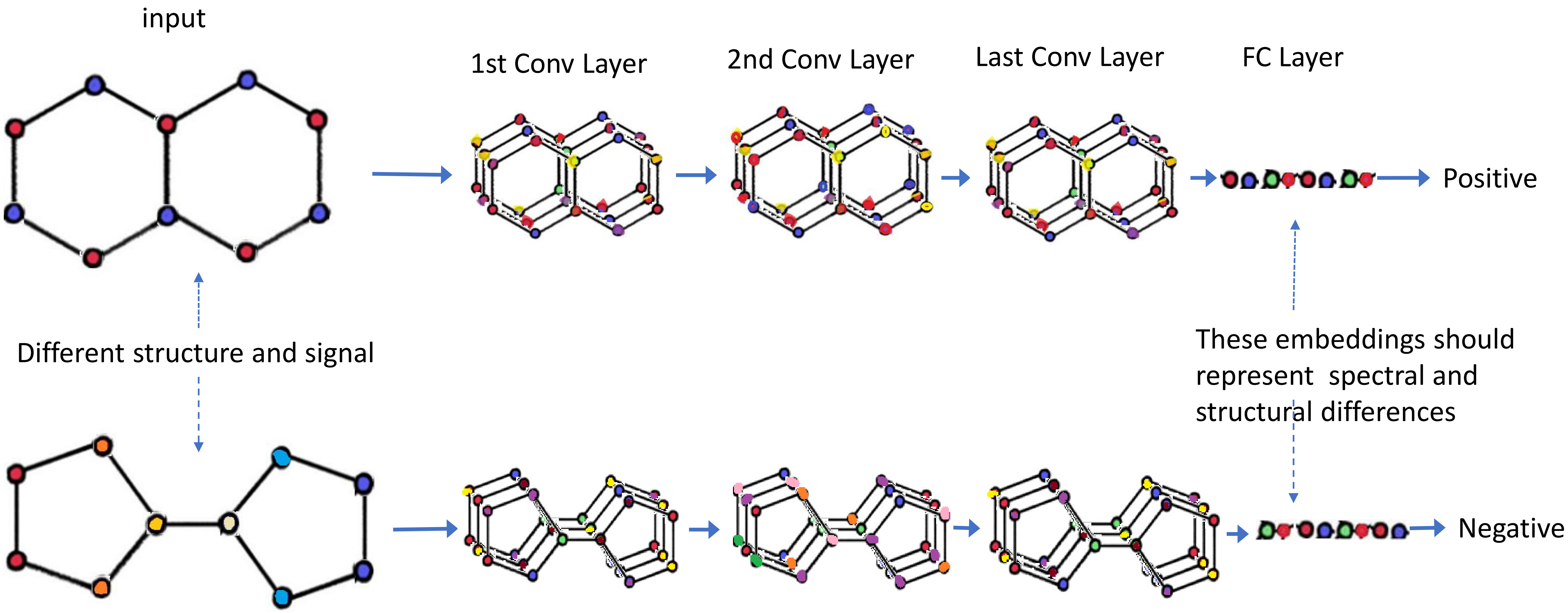


How CNN makes decision



Spectral features of image is the reason why CNN decides the image is cat or dog,
Not the structure where the signal lies on.

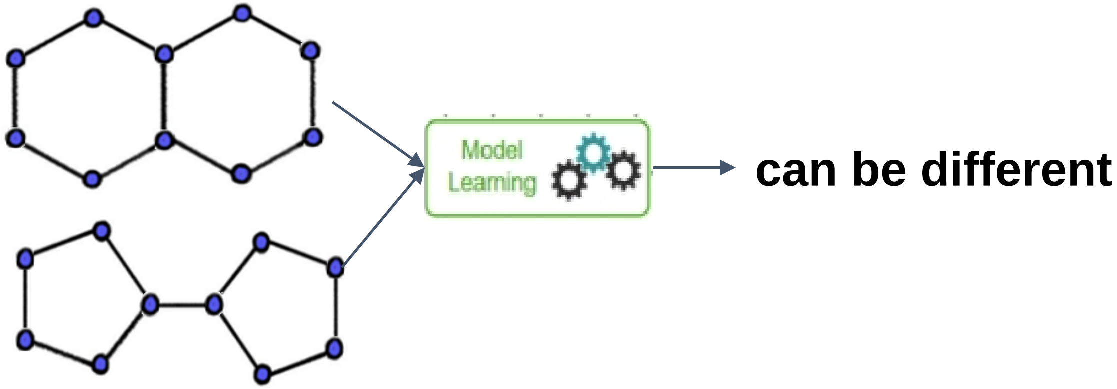
How GNN makes decision



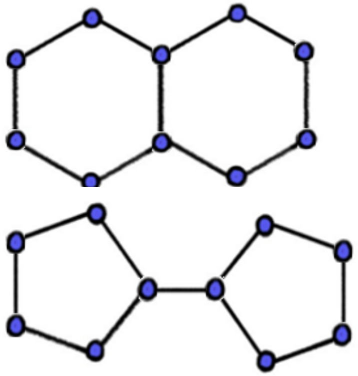
Spectral features of the graph signal and the structure of the graphs might be the reason why GNN decides it is positive or negative.

Structural Awareness

- Powerful structural aware GNN can
  - Embed all different graphs to the different latent points (isomorphism)



  - Count number of different substructure (triangle, star, 4-cycles, 5-cycles..etc)



Star	Triangle	4-cycle	5-cycle	6-cycle
2	0	0	0	2
2	0	0	2	0

(Structural) Expressive Power of GNN

- Measured by equivalence WL test order
  - 1-WL=2-WL <3-WL<4-WL<......<k-WL
- The higher k order, the more expressive power GNN has.
- WL power is also related to the ability on counting substructure.

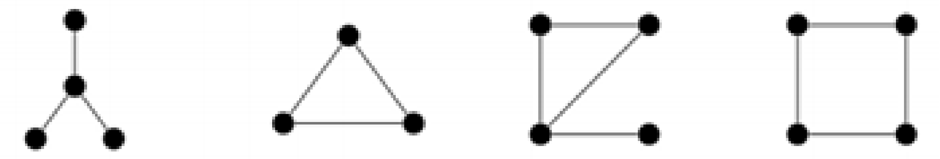


Figure 2. Sample of patterns: 3-star, triangle, tailed triangle and 4-cycle graphlets used in our analysis.

Theorem 3. *3-star graphlets can be counted by sentences in* $\mathcal{L}_1^+$.

Theorem 4. *Triangle and 4-cycle graphlets can be counted by sentences in* $\mathcal{L}_2^+$.

Theorem 5. *Tailed triangle graphlets can be counted by sentences in* $\mathcal{L}_3^+$.

From « Breaking the limits of Message Passing Graph Neural Networks »

(Structural) Expressive Power of GNN

- The higher k order, the more expressive power GNN has.
- k>2, k-WL GNN needs
  - O(n^(k-1)) memory, O(n^k) CPU time
- Naïve MPNN has 1-WL power and linear complexity.
- PPGN is one of the best 3-WL equivalent GNN which has quadratic memory cubic cpu complexity.

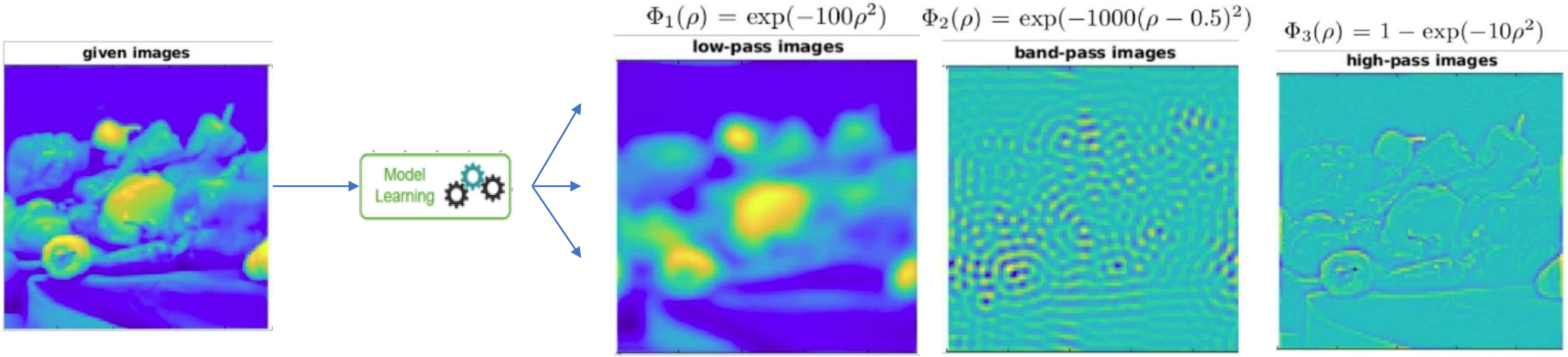
Theorem 1. *MPNNs such as GCN, GAT, GraphSage, GIN (defined in Appendix H) cannot go further than operations in $\mathcal{L}_1^+$. Thus, they are not more powerful than the 1-WL test.*

Theorem 2. *Chebnet is more powerful than the 1-WL test if the Laplacian maximum eigenvalues of the non-regular graphs to be compared are not the same. Otherwise Chebnet is not more powerful than 1-WL.*

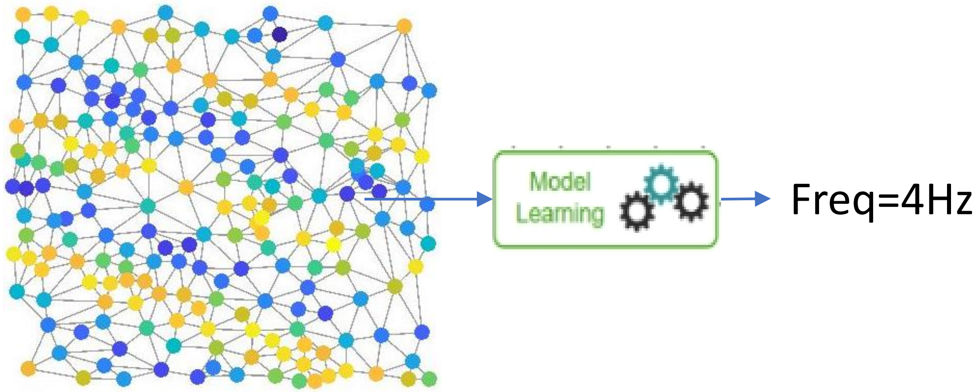
From « Breaking the limits of Message Passing Graph Neural Networks »

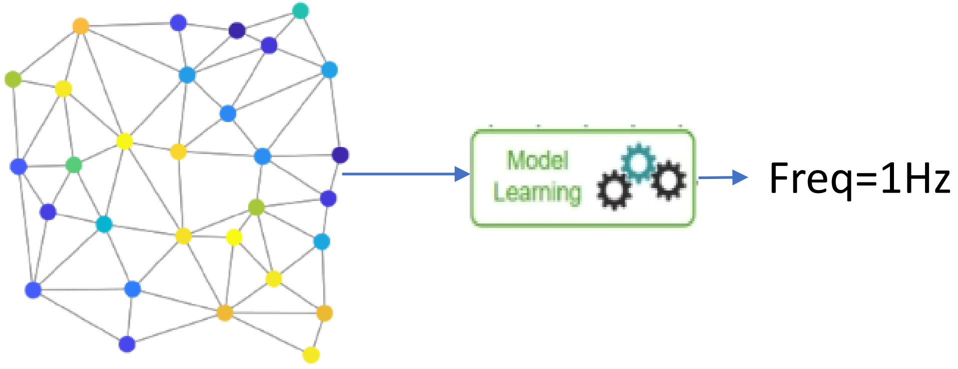
Spectral Awareness

- Powerful spectral aware GNN can learn all kind of filtering effects.



- It should aware of spectrum of the signal on the graph





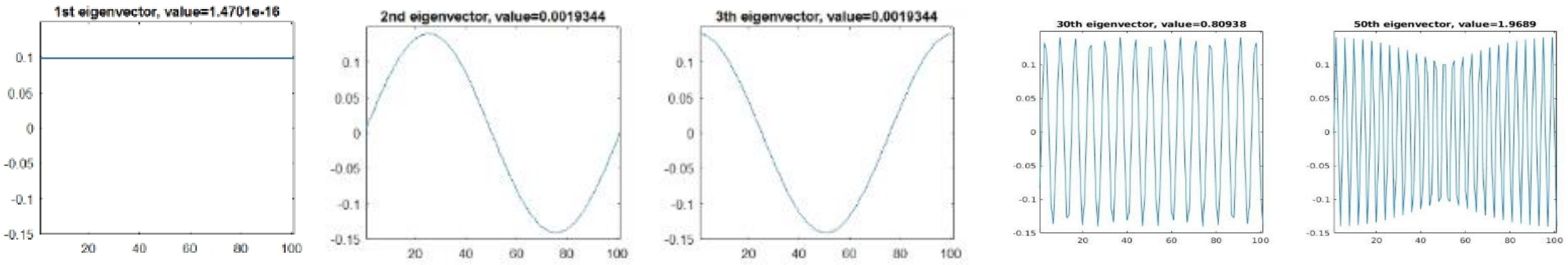
Graph Signal Processing

Laplacian of Graph $L = I - D^{-1/2} A D^{-1/2}$

Eigenvector and eigenvalues

$$U^{-1} L U = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$$

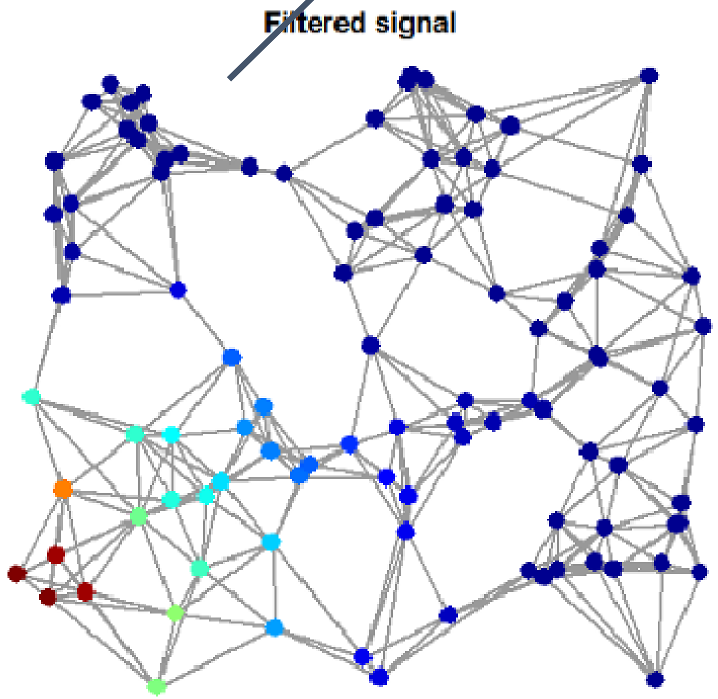
$$0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$$

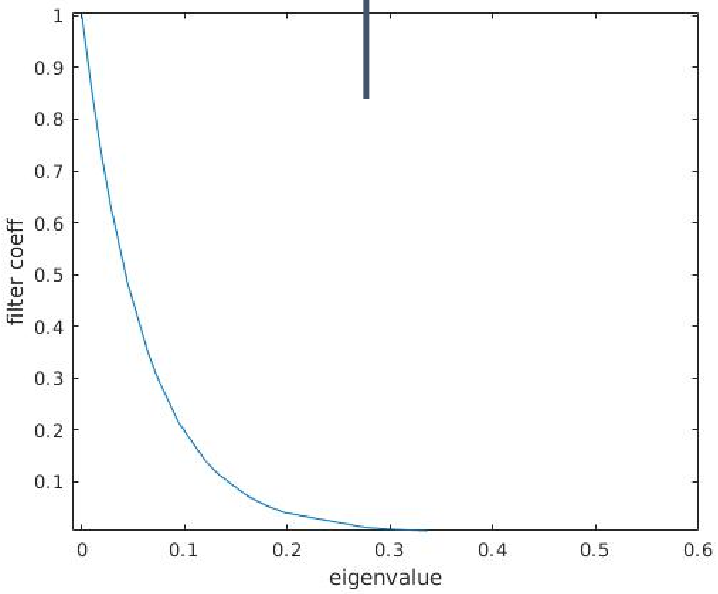


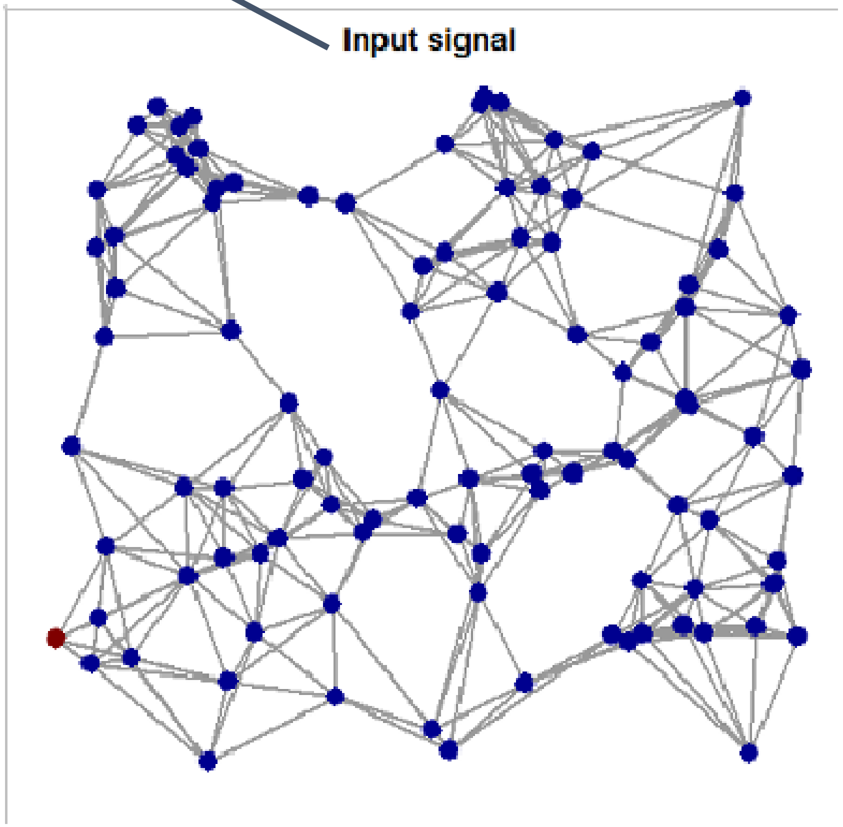
Any signal can be written by weighted sum of these base functions

Spectral Graph Filtering

$$x_{filtered} = U diag(F(\boldsymbol{\lambda}))U^{\top}x,$$







Message Passing (Graph) Neural Network

- In our research[1], we generalize spatial and spectral GNN by

$$H^{(l+1)} = \sigma\Big(\sum_s C^{(s)} H^{(l)} W^{(l,s)}\Big),$$

Convolution Support　　Node Features　　Trainable Parameters

- Spatial Methods are defined by C matrices, Such $C^{(0)}$=A, $C^{(1)}$=I,…
- Spectral Method defined by $\Phi_j(\lambda_i)$ such as $\Phi_s(\lambda) = exp(-b(\lambda - f_s)^2)$,
- Where transition can be written by $C^{(s)} = U \,\mathrm{diag}(\Phi_s(\boldsymbol{\lambda}))U^\top.$

1 Balcilar et al. Analyzing the expressive power of graph neural networks in a spectral perspective. ICLR2021.

Bridging the Gap Between Spectral and Spatial MPNN

- Spatial to Spectral transition

Corollary 1.1. *The frequency profile* $\Phi_s(\boldsymbol{\lambda}) = diag^{-1}(U^\top C^{(s)} U)$.

- Definition of Spatial & Spectral MPNN

Definition 2. *A **Spectral-designed** graph convolution refers to a convolution where supports are written as a function of eigenvalues ($\Phi_s(\boldsymbol{\lambda})$) and eigenvectors ($U$) of the corresponding graph Laplacian (equation $\boxed{6}$). Thus, each convolution support $C^{(s)}$ has the same frequency response $\Phi_s(\boldsymbol{\lambda})$ over different graphs. Graph convolution out of this definition is called **spatial-designed** graph convolution.*

Spectral Analysis of some MPNNs

Theorem 2. *The theoretical frequency response of each support of ChebNet can be defined as*

$$\Phi_1(\boldsymbol{\lambda}) = \mathbf{1}, \;\; \Phi_2(\boldsymbol{\lambda}) = \frac{2\boldsymbol{\lambda}}{\lambda_{\max}} - \mathbf{1}, \;\; \Phi_k(\boldsymbol{\lambda}) = 2\Phi_2(\boldsymbol{\lambda})\Phi_{k-1}(\boldsymbol{\lambda}) - \Phi_{k-2}(\boldsymbol{\lambda}),$$

where $\mathbf{1}$ *is the vector of ones and* $\lambda_{\max}$ *is the maximum eigenvalue.*

Theorem 3. *The theoretical frequency response of each support of CayleyNet can be defined as*

$$\Phi_s(\boldsymbol{\lambda}) = \begin{cases} \mathbf{1} & \textit{if } s = 1 \\ \cos(\frac{s}{2}\theta(h\boldsymbol{\lambda})) & \textit{if } s \in \{2, 4, \ldots, 2r\} \\ -\sin(\frac{s-1}{2}\theta(h\boldsymbol{\lambda})) & \textit{if } s \in \{3, 5, \ldots, 2r+1\} \end{cases}$$

where h *is a trainable scalar and* $\theta(x) = atan2(-1, x) - atan2(1, x)$.

Spectral Analysis of some MPNNs

Theorem 4. *The theoretical frequency response of GCN support can be approximated as*

$$\Phi(\boldsymbol{\lambda}) \approx \mathbf{1} - \boldsymbol{\lambda}\overline{p}/(\overline{p}+1),$$

where $\overline{p}$ is the average node degree in the graph.

Theorem 5. *The theoretical frequency response of GIN support can be approximted as*

$$\Phi(\boldsymbol{\lambda}) \approx \overline{p}\left(\frac{1+\epsilon}{\overline{p}} + \mathbf{1} - \boldsymbol{\lambda}\right)$$

where ϵ is a trainable scalar.

Table 1: Summary of the studied GNN models.

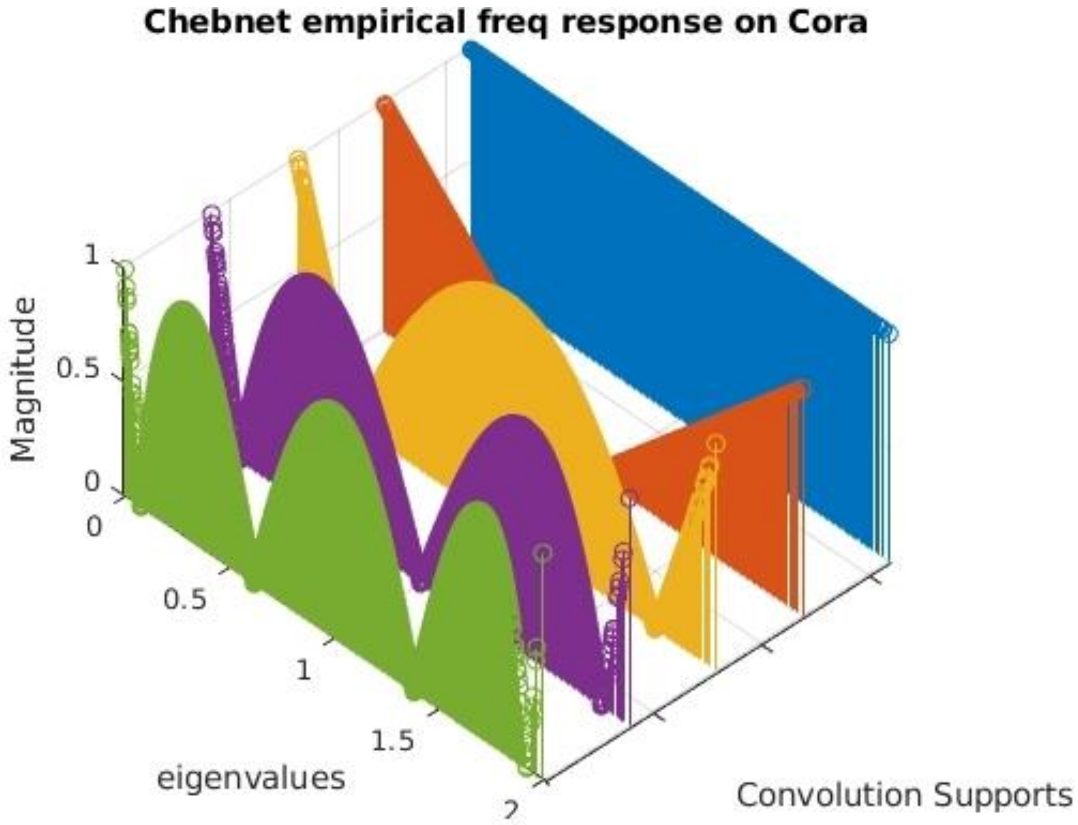
	Design	Support Type	Convolution Matrix	Frequency Response
MLP	Spectral	Fixed	$C = I$	$\Phi(\boldsymbol{\lambda}) = \mathbf{1}$
GCN	Spatial	Fixed	$C = \tilde{D}^{-0.5}\tilde{A}\tilde{D}^{-0.5}$	$\Phi(\boldsymbol{\lambda}) \approx \mathbf{1} - \boldsymbol{\lambda}\overline{p}/(\overline{p}+1)$
GIN	Spatial	Trainable	$C = A + (1+\epsilon)I$	$\Phi(\boldsymbol{\lambda}) \approx \overline{p}\left(\frac{1+\epsilon}{\overline{p}} + \mathbf{1} - \boldsymbol{\lambda}\right)$
GAT	Spatial	Trainable	$C^{(s)}_{v,u} = e_{v,u}/\sum_{k\in\tilde{\mathcal{N}}(v)} e_{v,k}$	NA
CayleyNet[a]	Spectral	Trainable	$C^{(1)} = I$ $C^{(2r)} = Re(\rho(hL)^r)$ $C^{(2r+1)} = Re(\mathbf{i}\rho(hL)^r)$	$\Phi_1(\boldsymbol{\lambda}) = \mathbf{1}$ $\Phi_{2r}(\boldsymbol{\lambda}) = \cos(r\theta(h\boldsymbol{\lambda}))$ $\Phi_{2r+1}(\boldsymbol{\lambda}) = -\sin(r\theta(h\boldsymbol{\lambda}))$
ChebNet	Spectral	Fixed	$C^{(1)} = I$ $C^{(2)} = 2L/\lambda_{\max} - I$ $C^{(s)} = 2C^{(2)}C^{(s-1)} - C^{(s-2)}$	$\Phi_1(\boldsymbol{\lambda}) = \mathbf{1}$ $\Phi_2(\boldsymbol{\lambda}) = 2\boldsymbol{\lambda}/\lambda_{\max} - \mathbf{1}$ $\Phi_s(\boldsymbol{\lambda}) = 2\Phi_2(\boldsymbol{\lambda})\Phi_{s-1}(\boldsymbol{\lambda}) - \Phi_{s-2}(\boldsymbol{\lambda})$

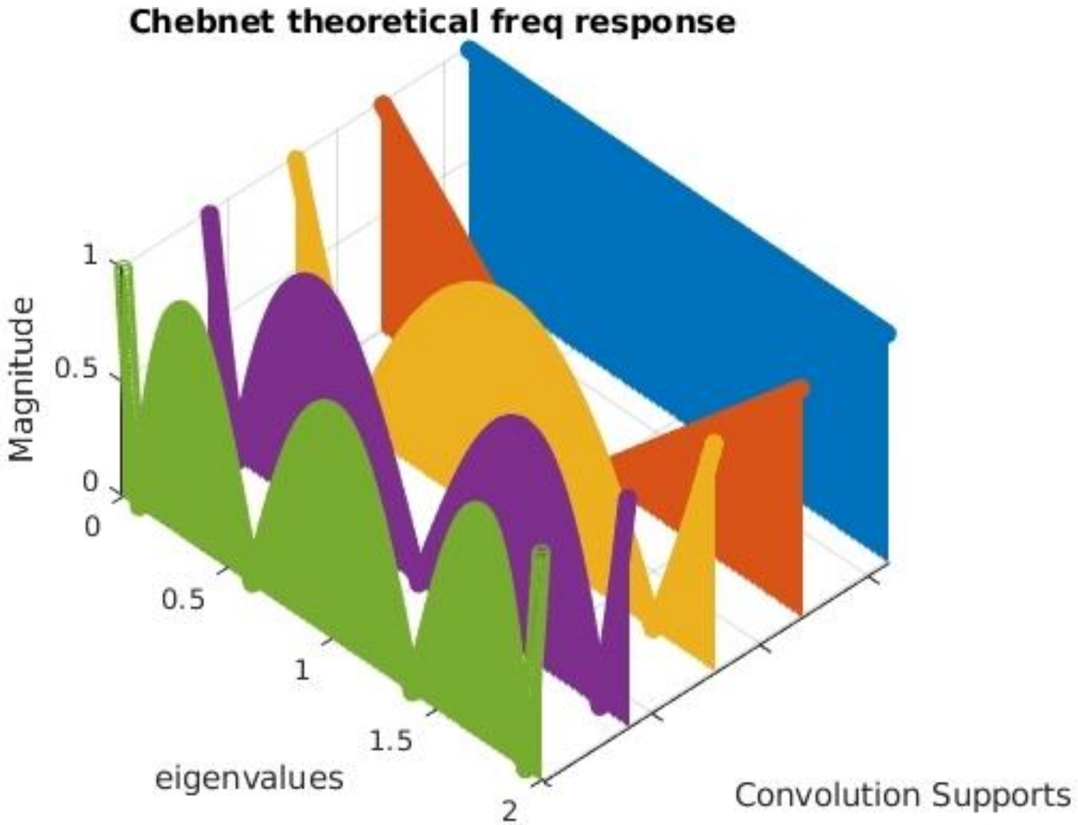
[a] $\rho(x) = (x - \mathbf{i}I)/(x + \mathbf{i}I)$

Chebnet, Spectral Designed, Fixed Support

$$C^{(1)} = I$$
$$C^{(2)} = 2L/\lambda_{\max} - I$$
$$C^{(s)} = 2C^{(2)}C^{(s-1)} - C^{(s-2)}$$

$$\Phi_1(\boldsymbol{\lambda}) = \mathbf{1}$$
$$\Phi_2(\boldsymbol{\lambda}) = 2\boldsymbol{\lambda}/\lambda_{\max} - \mathbf{1}$$
$$\Phi_s(\boldsymbol{\lambda}) = 2\Phi_2(\boldsymbol{\lambda})\Phi_{s-1}(\boldsymbol{\lambda}) - \Phi_{s-2}(\boldsymbol{\lambda})$$

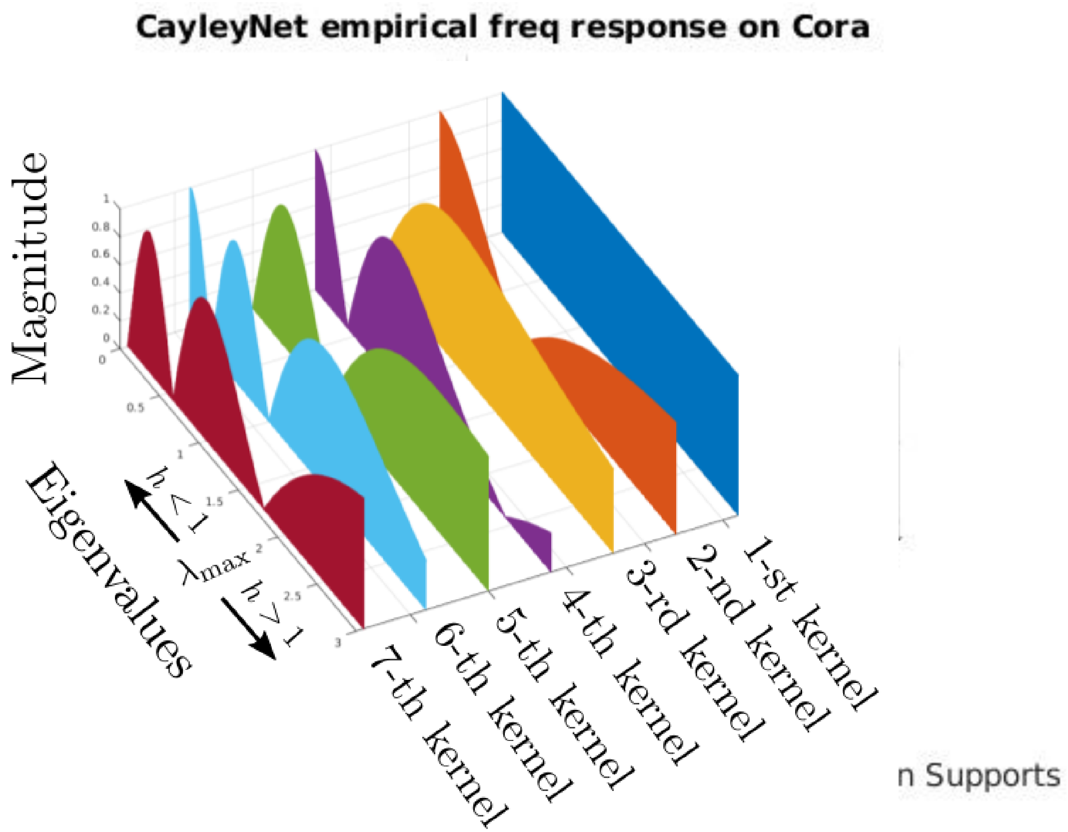


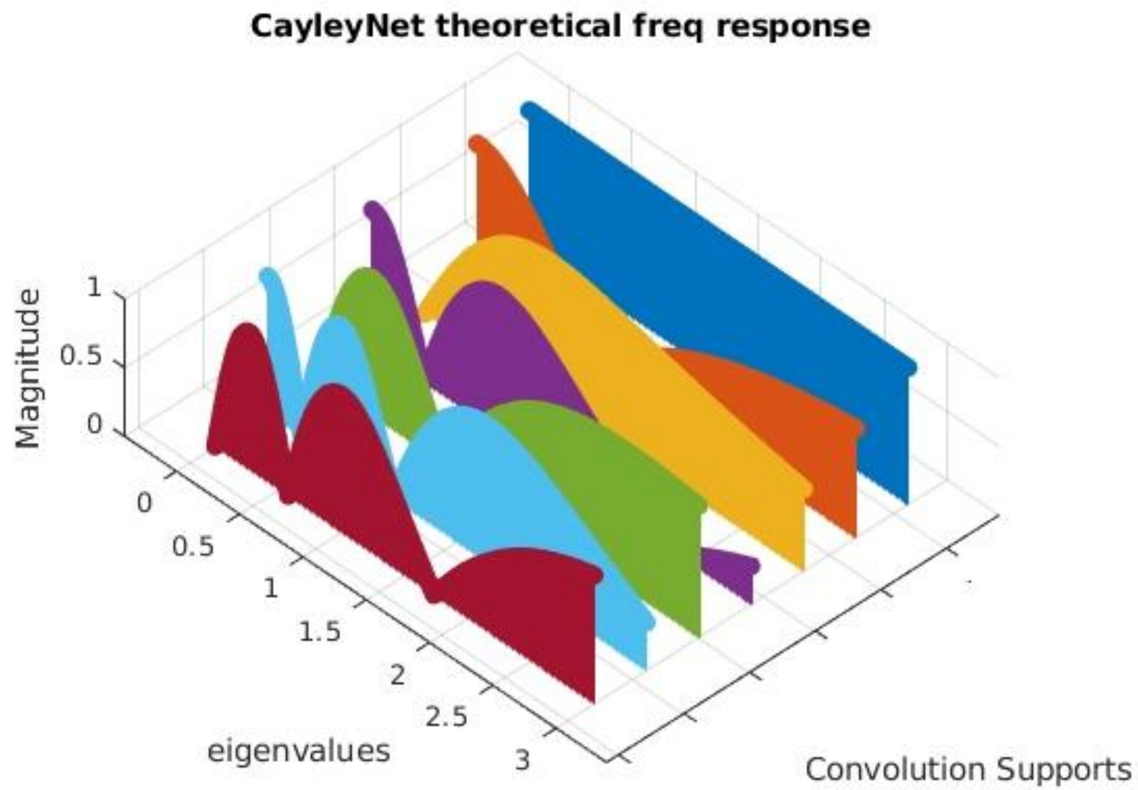


CayleyNet Spectral Designed, Trainable Support

$$C^{(1)} = I$$
$$C^{(2r)} = Re(\rho(hL)^r)$$
$$C^{(2r+1)} = Re(\mathbf{i}\rho(hL)^r)$$

$$\Phi_1(\boldsymbol{\lambda}) = \mathbf{1}$$
$$\Phi_{2r}(\boldsymbol{\lambda}) = \cos(r\theta(h\boldsymbol{\lambda}))$$
$$\Phi_{2r+1}(\boldsymbol{\lambda}) = -\sin(r\theta(h\boldsymbol{\lambda}))$$

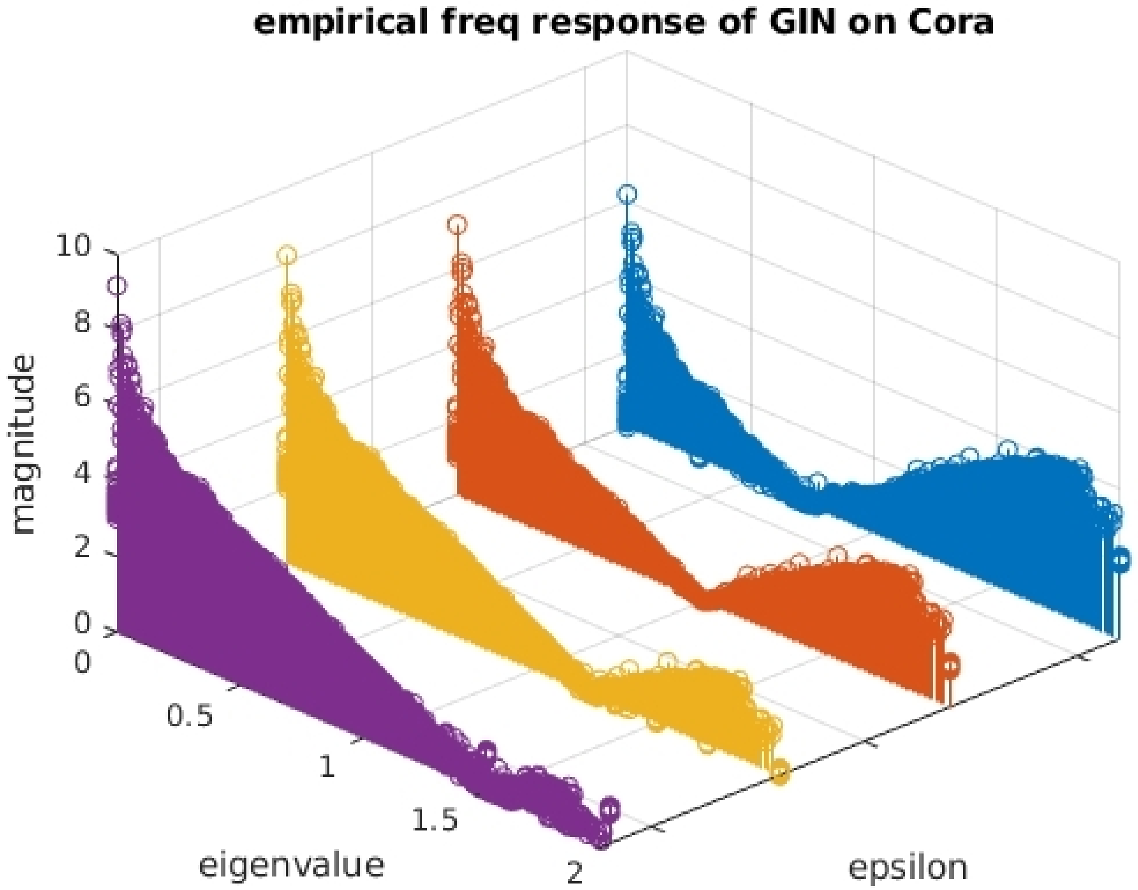


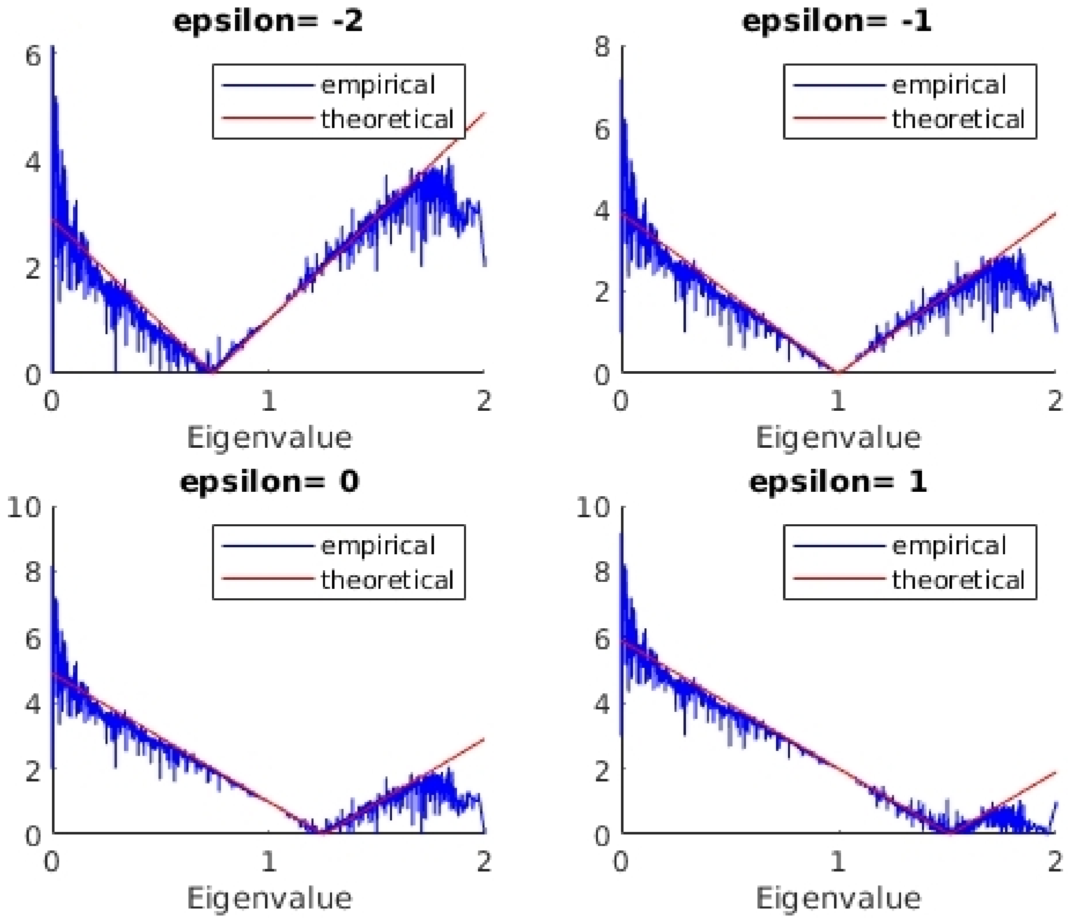


GIN Spatial Designed, Trainable Support

$$C = A + (1+\epsilon)I$$

$$\Phi(\boldsymbol{\lambda}) \approx \overline{p}\left(\frac{1+\epsilon}{\overline{p}} + \mathbf{1} - \boldsymbol{\lambda}\right)$$

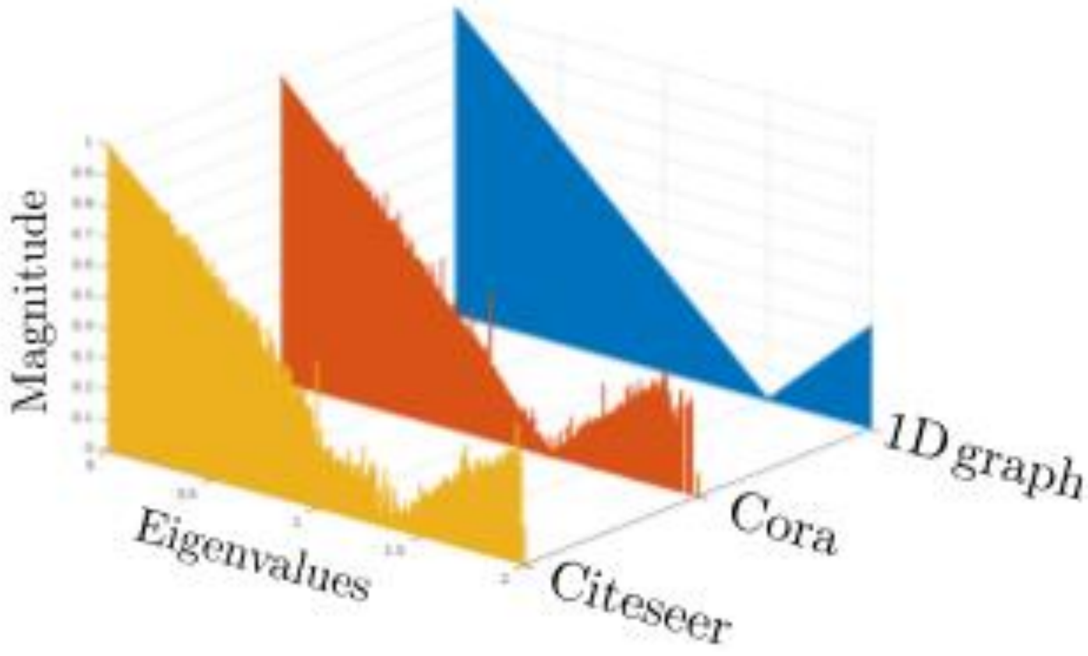




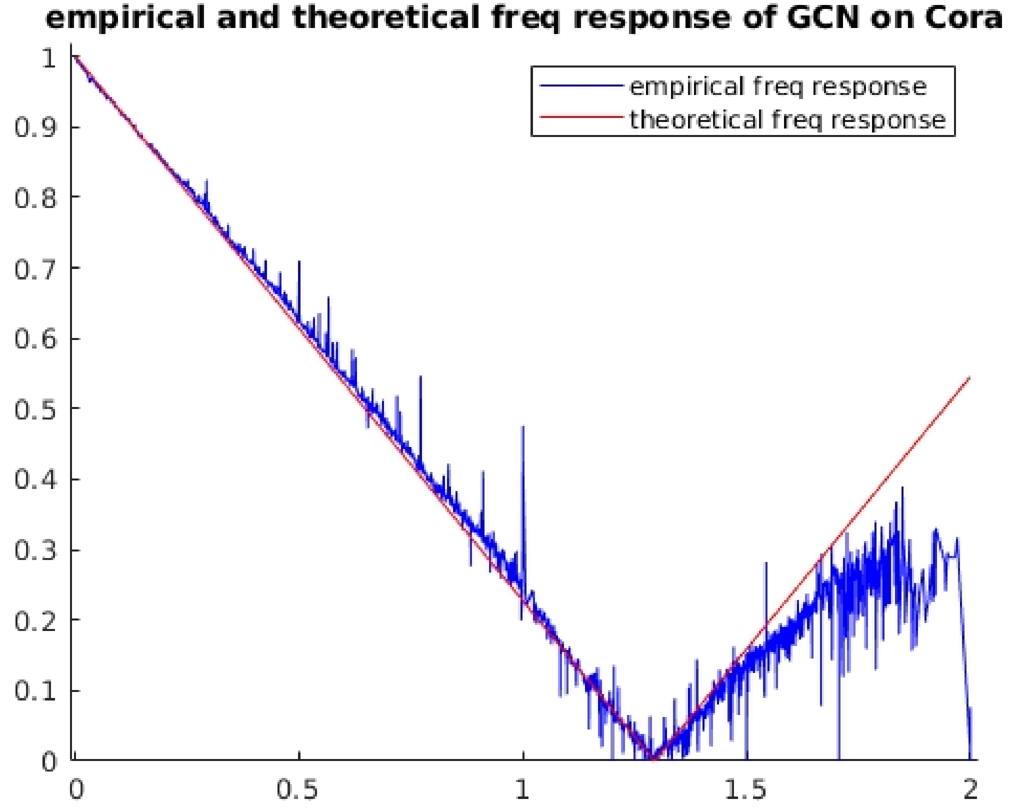
GCN Spatial Designed, Fixed Support

$$C = \tilde{D}^{-0.5}\tilde{A}\tilde{D}^{-0.5}$$

$$\Phi(\boldsymbol{\lambda}) \approx \mathbf{1} - \boldsymbol{\lambda}\overline{p}/(\overline{p}+1)$$

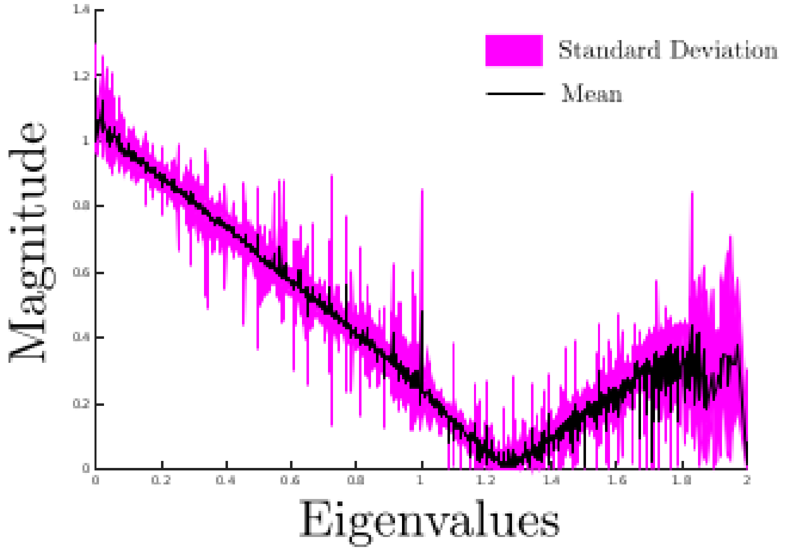


(a) GCN frequency profiles

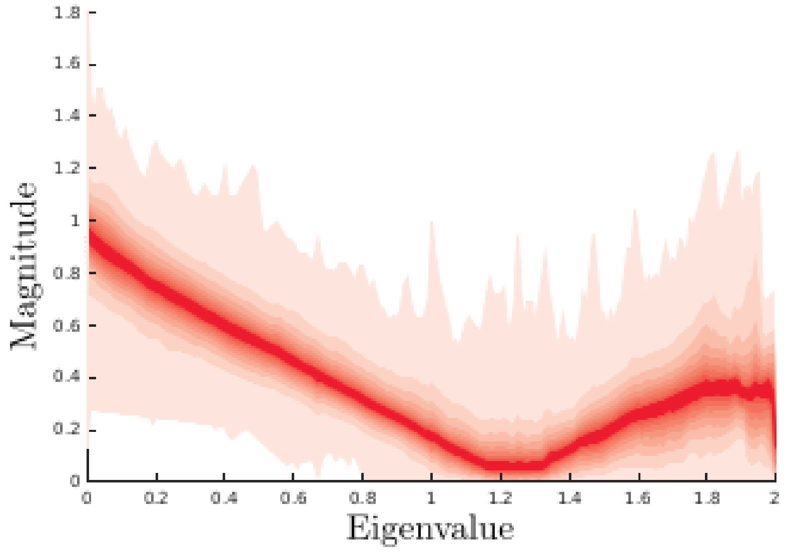


GAT Spatial Designed, Trainable Support

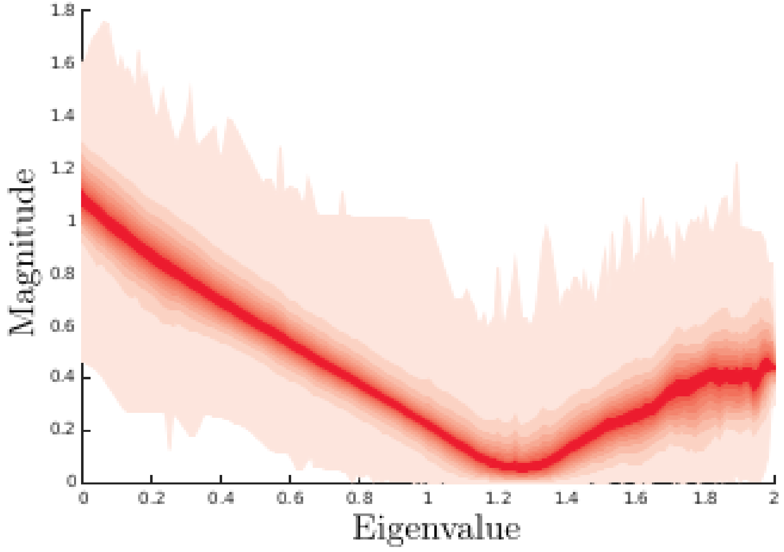
$$C_{v,u}^{(s)} = e_{v,u} / \sum_{k \in \tilde{N}(v)} e_{v,k}$$



(a) Expected frequency response from Simulation on Cora



(b) Heat density map of learned frequency response on ENZYMES



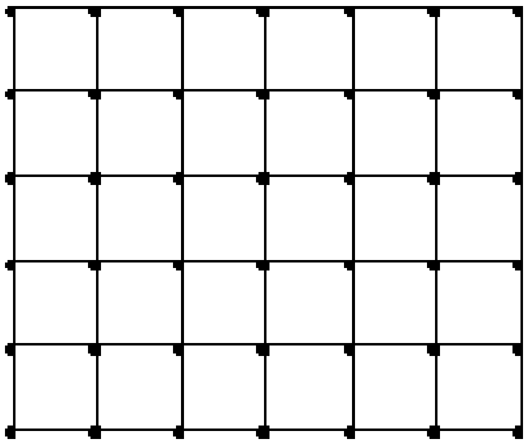
(c) Heat density map of learned frequency response on PROTEINS

Spectral Expressive Power of GNN

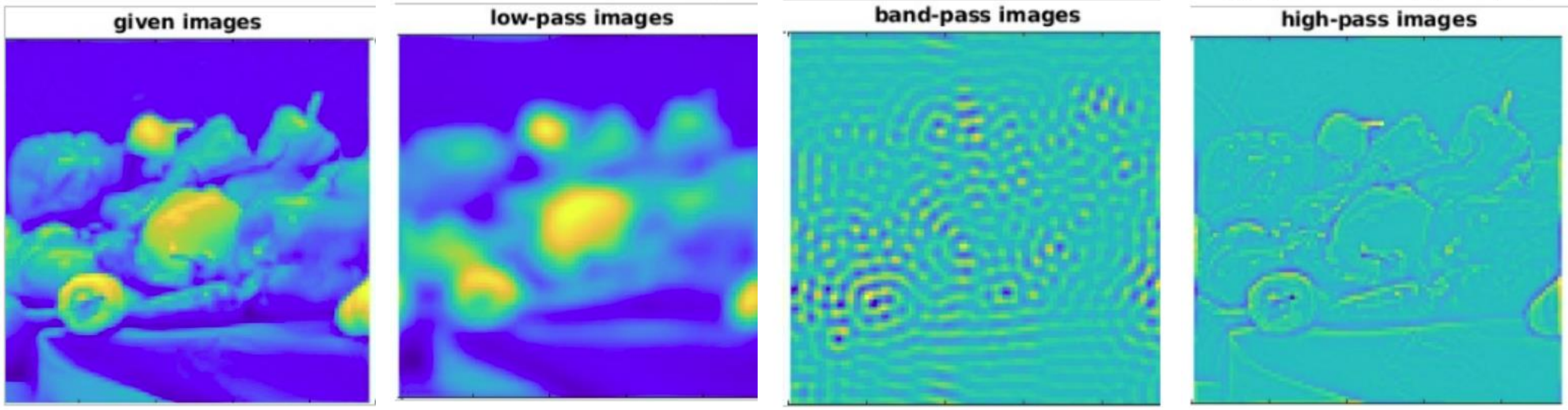
- We do not propose some metric to measure spectral expressive power.
- But just present what the frequency responses are.
- As conclusion of Frequency Responses:
    - Spatial MPNN is nothing but just low-pass filter!
    - Spectral MPNN cover the spectrum well but not have band specific filters
    - Most of the natural graph problems need low-pass effects.
    - If the signal on graph matters, spectral methods are best!

Experiments

- 2DGrid graph

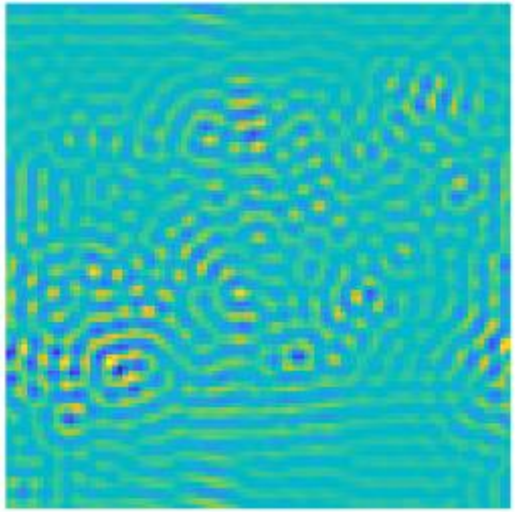


$\Phi_1(\rho) = \exp(-100\rho^2)$ $\Phi_2(\rho) = \exp(-1000(\rho - 0.5)^2)$ $\Phi_3(\rho) = 1 - \exp(-10\rho^2)$



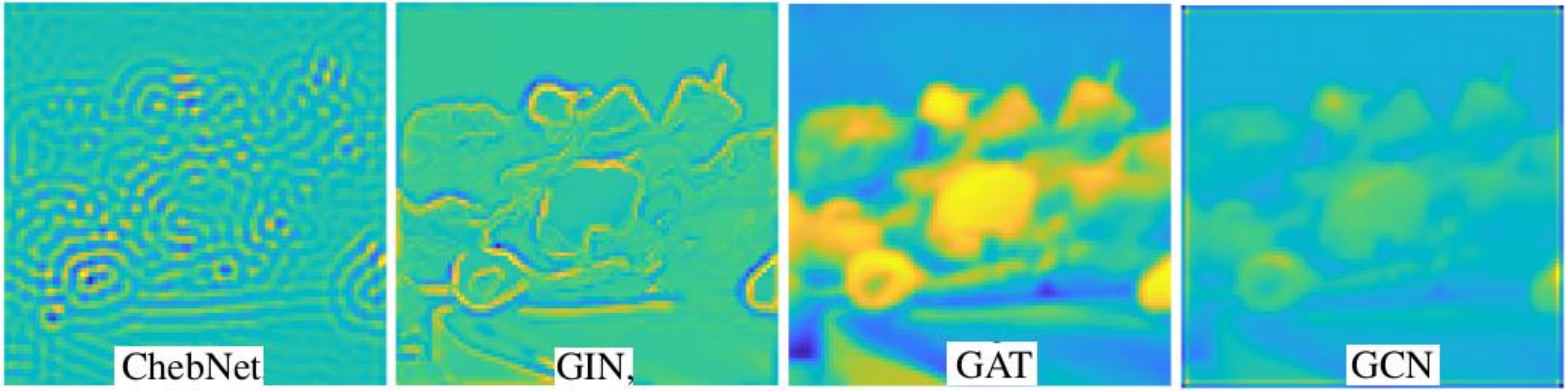
Experiments

Ground truth



Prediction Target	GCN	GIN	GAT	ChebNet
Low-pass filter (Φ_1)	15.55	11.01	10.50	3.44
Band-pass filter (Φ_2)	79.72	63.24	79.68	17.30
High-pass filter (Φ_3)	29.51	14.27	29.10	2.04

Prediction

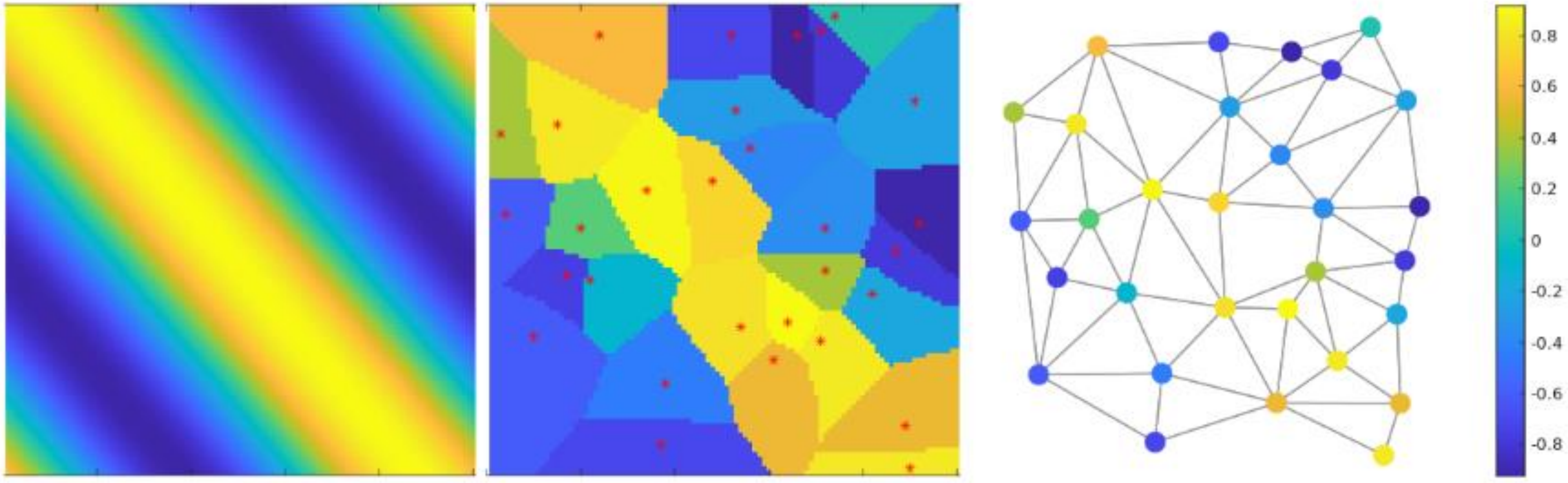


Experiments

- BandClass

Table 5: Test set accuracy and binary cross entropy loss.

	MLP	GCN	GIN	GAT	ChebNet
Accuracy	50	77.90	87.60	85.30	98.2
Loss	0.69	0.454	0.273	0.324	0.062



Experiments

- MNIST-75

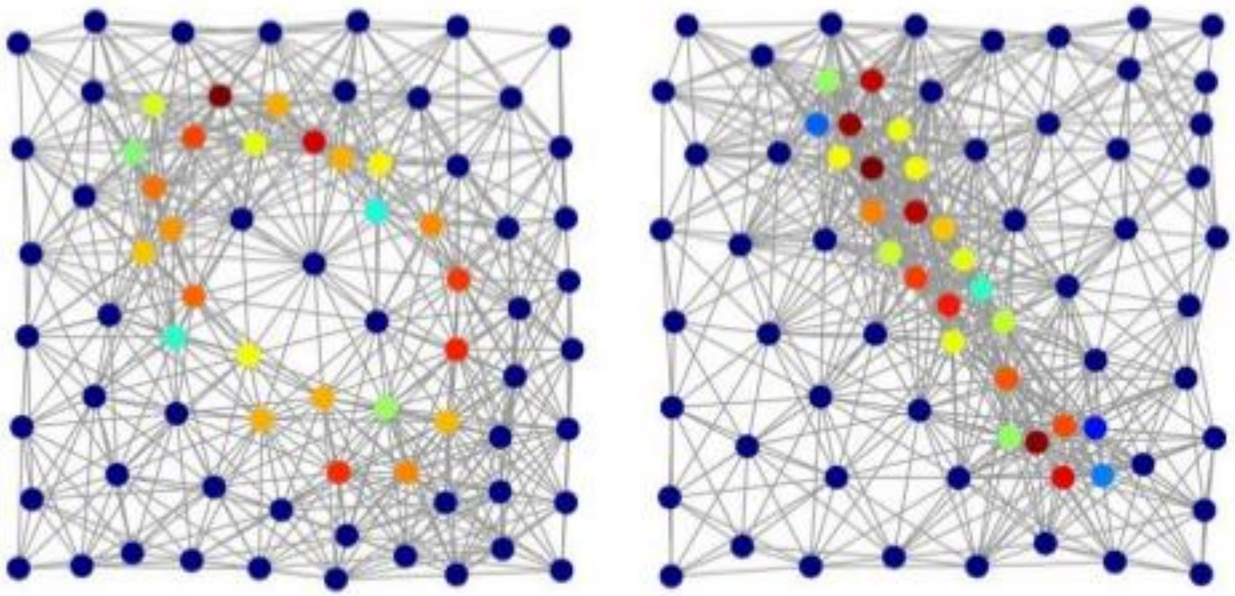


Table 2: Test set accuracies on MNIST superpixel dataset

Node feature	MLP	GCN	GIN	GAT	CayleyNet	ChebNet
Node degree	11.29±0.5	15.81±0.8	32.45±1.2	31.72±1.5	45.61±1.7	46.23±1.8
Pixel value	12.11±0.5	11.35±1.1	64.96±3.9	62.61±2.9	88.41±2.1	91.10±1.9
Both	25.10±1.2	52.98±3.1	75.23±4.1	82.73±2.1	90.31±2.3	92.08±2.2

Summary of Our Contributions

- Bridging the gap between Spatial-Spectral MPNN
- Show how to do spectral analysis of GNN.
- Show spatial MPNN is nothing but low-pass filter.
- Propose new taxonomy on GNN.
- Put a new criteria on theoretical evaluation of expressive power of GNN.

https://github.com/balcilar/gnn-spectral-expressive-power

Back to the title: Which is more important?

- Low-pass GNN results are comparable on Assortative graph dataset.
- Low-pass GNN results are poor on Disassortative dataset.
- So far, I did not find natural graph dataset that low-pass GNN fails ridiculously compare to the spectrally powerful method.
- But there are many dataset that structural awareness is more important.
- Since we cannot know which is important in advance, we should be aware of both.

MODEL	ZINC12K	MNIST-75
MLP	0.5869 ± 0.025	$25.10\% \pm 0.12$
GCN	0.3322 ± 0.010	$52.80\% \pm 0.31$
GAT	0.3977 ± 0.007	$82.73\% \pm 0.21$
GIN	0.3044 ± 0.010	$75.23\% \pm 0.41$
CHEBNET	0.3569 ± 0.012	$\mathbf{92.08\% \pm 0.22}$
PPGN	$\mathbf{0.1589 \pm 0.007}$	$90.04\% \pm 0.54$
GNNML1	0.3140 ± 0.015	$84.21\% \pm 1.75$
GNNML3	$\mathbf{0.1612 \pm 0.006}$	$\mathbf{91.98\% \pm 0.18}$