Data-Driven Learning of Geometric Scattering Networks

Alexander Tong*, Frederik Wenkel*, Kincaid MacDonald, Smita Krishnaswamy, Guy Wolf

Presented at MLSP 2021

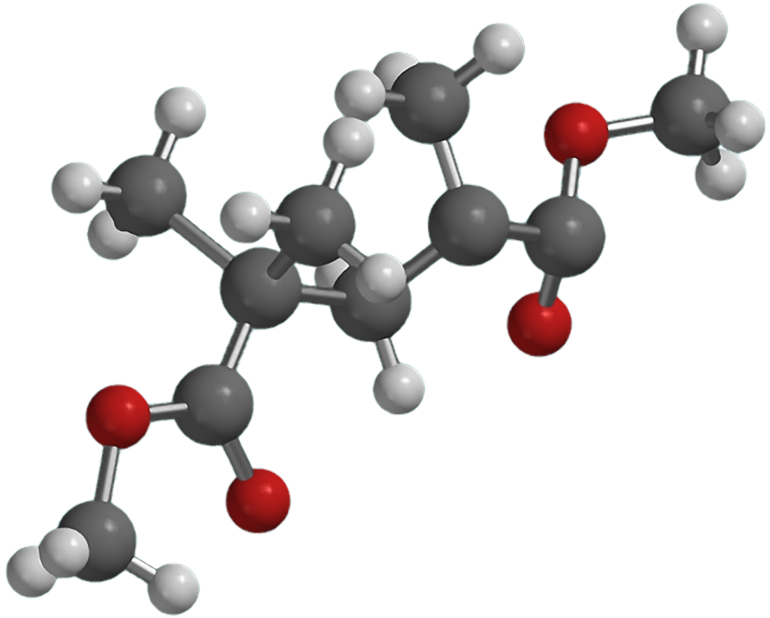
* Equal Contribution

Geometric Deep Learning

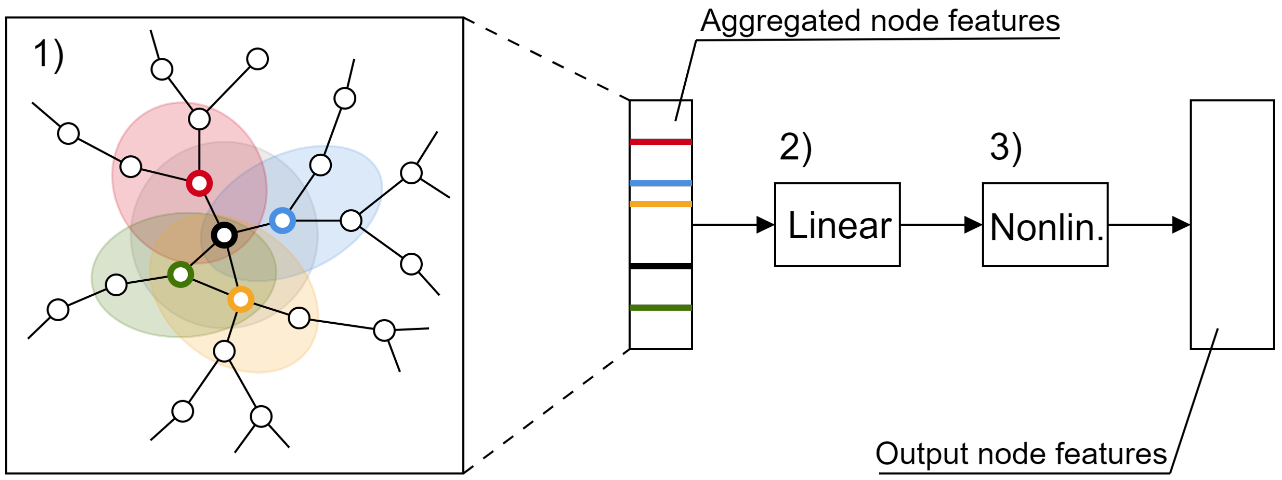
Goal: Generalize networks operating on Euclidean structures to non-Euclidean geometries.

Graphs are a natural model for many datatypes:

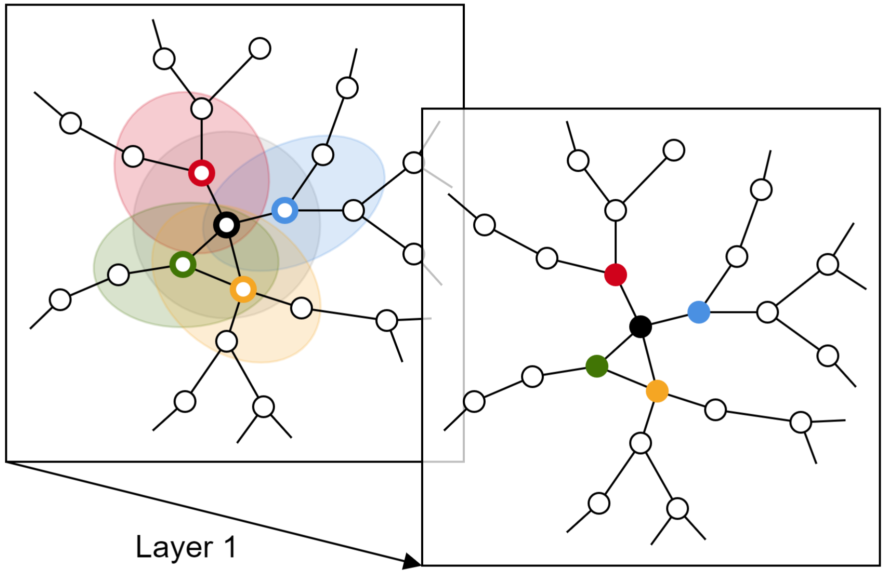
- Citation networks
- Social networks
- Molecule structures



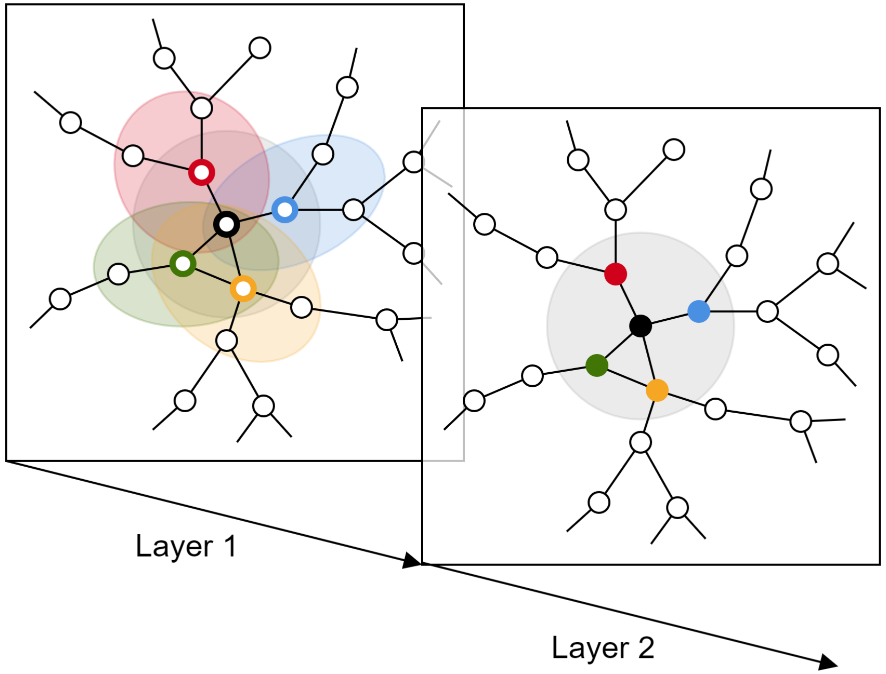
Graph Convolutional Networks [Kipf and Welling 2016]



Each layer aggregates information and can be stacked for larger filters



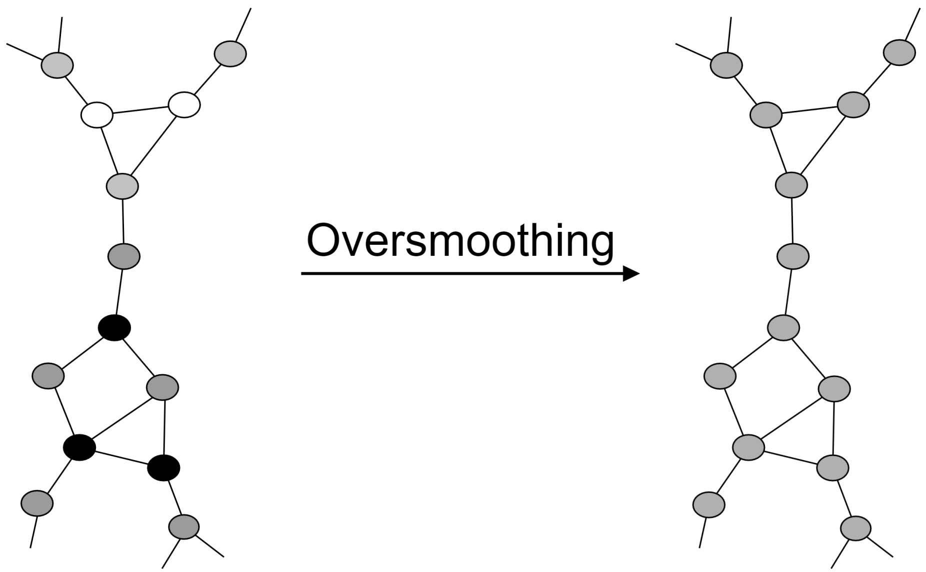
Each layer aggregates information and can be stacked for larger filters



… at a cost

"Oversmoothing" [Li et al. 2018]

GCN has inductive bias towards averaging features limiting the use of additional depth

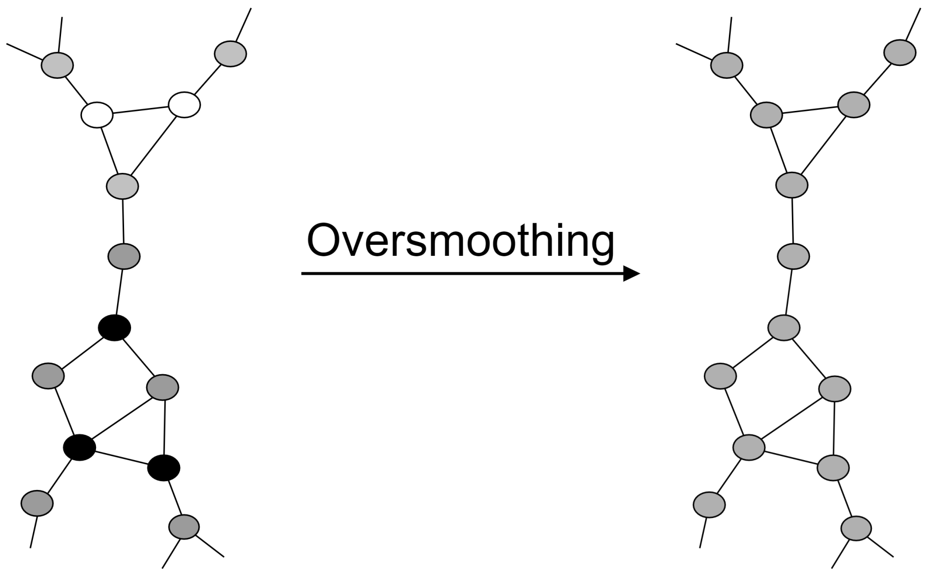


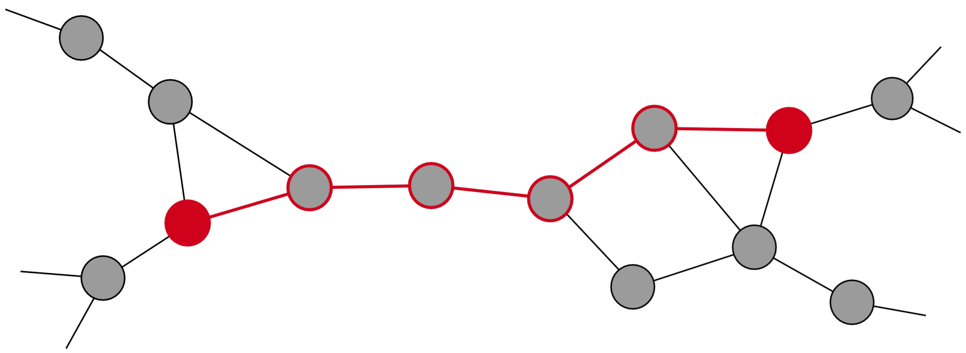
… at a cost

"Oversmoothing" [Li et al. 2018]

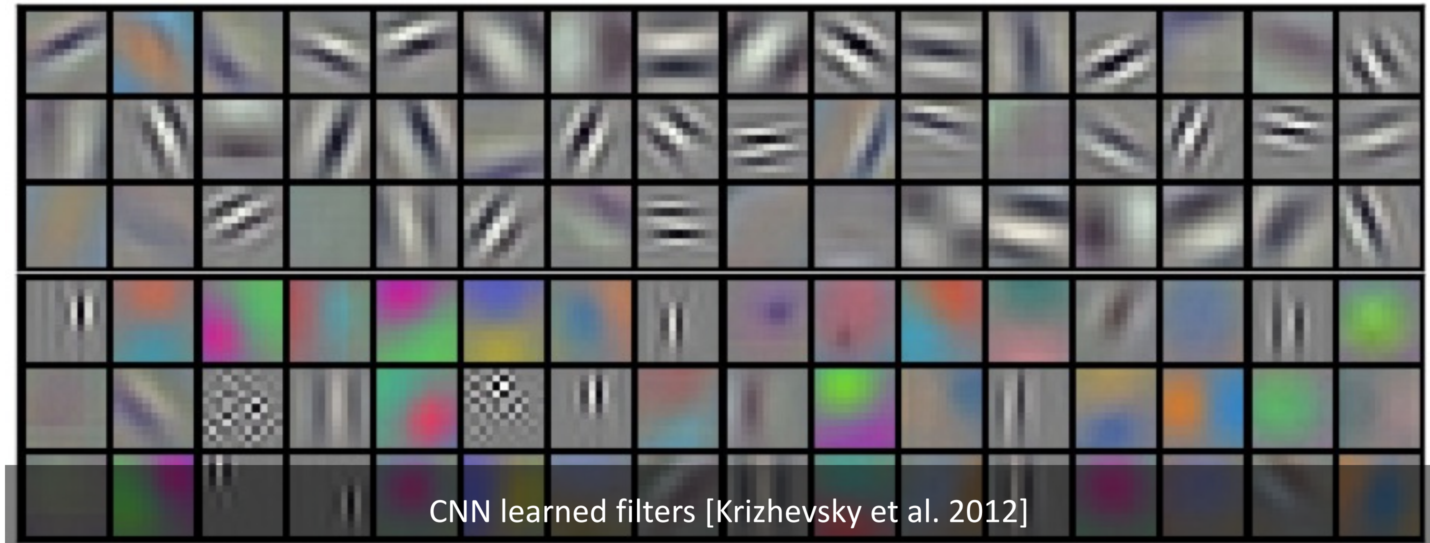
"Underreaching" [Barcelo et al. 2020]

GCN can only aggregate information at distance equal to the number of layers

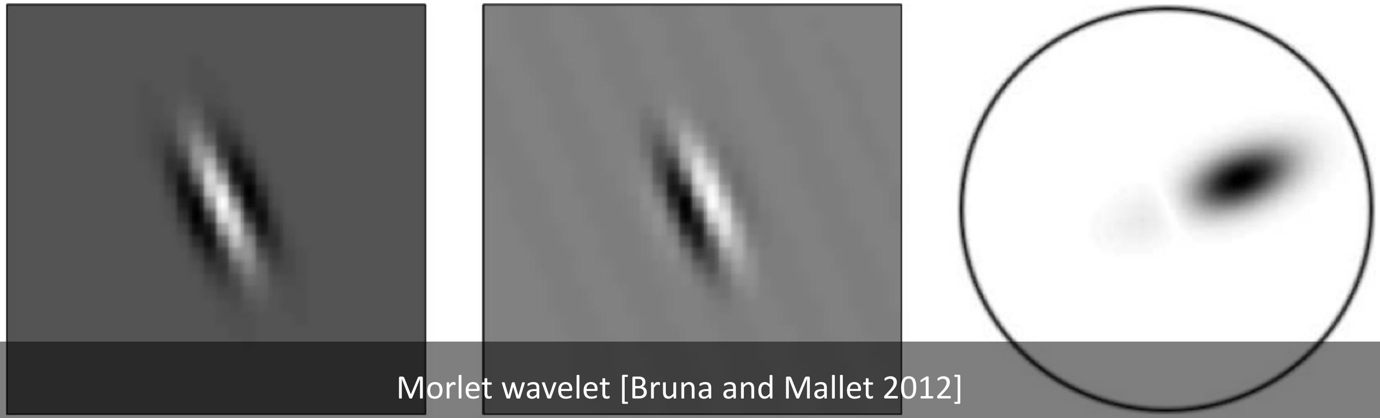




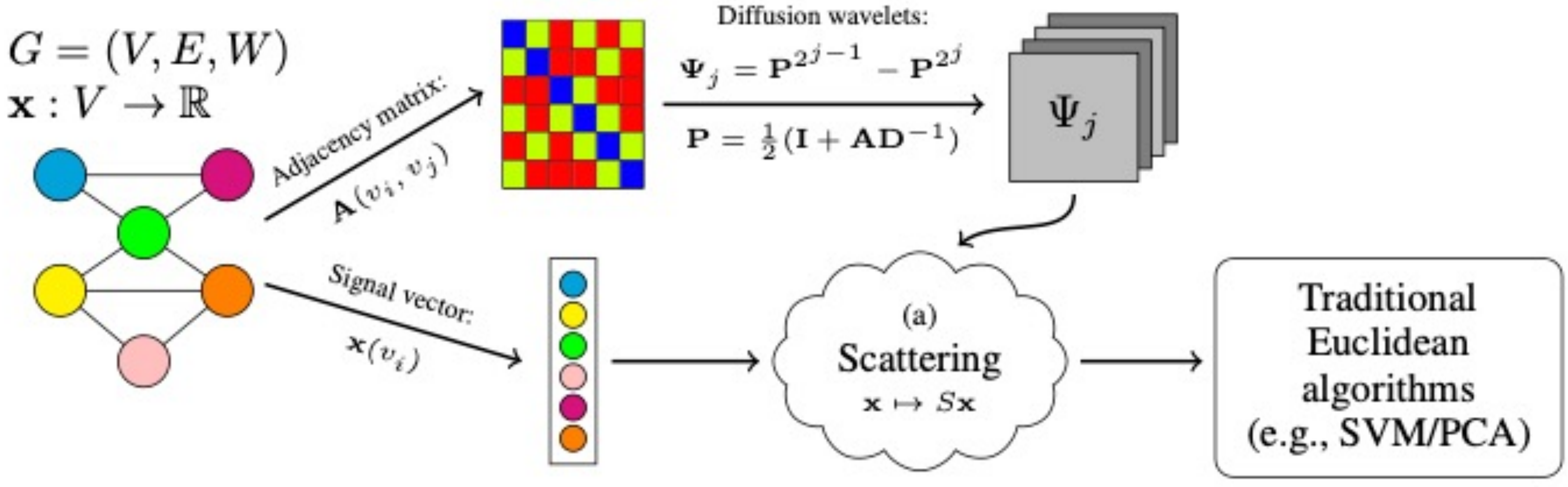
(Euclidean) Scattering



CNN learned filters [Krizhevsky et al. 2012]



Morlet wavelet [Bruna and Mallet 2012]

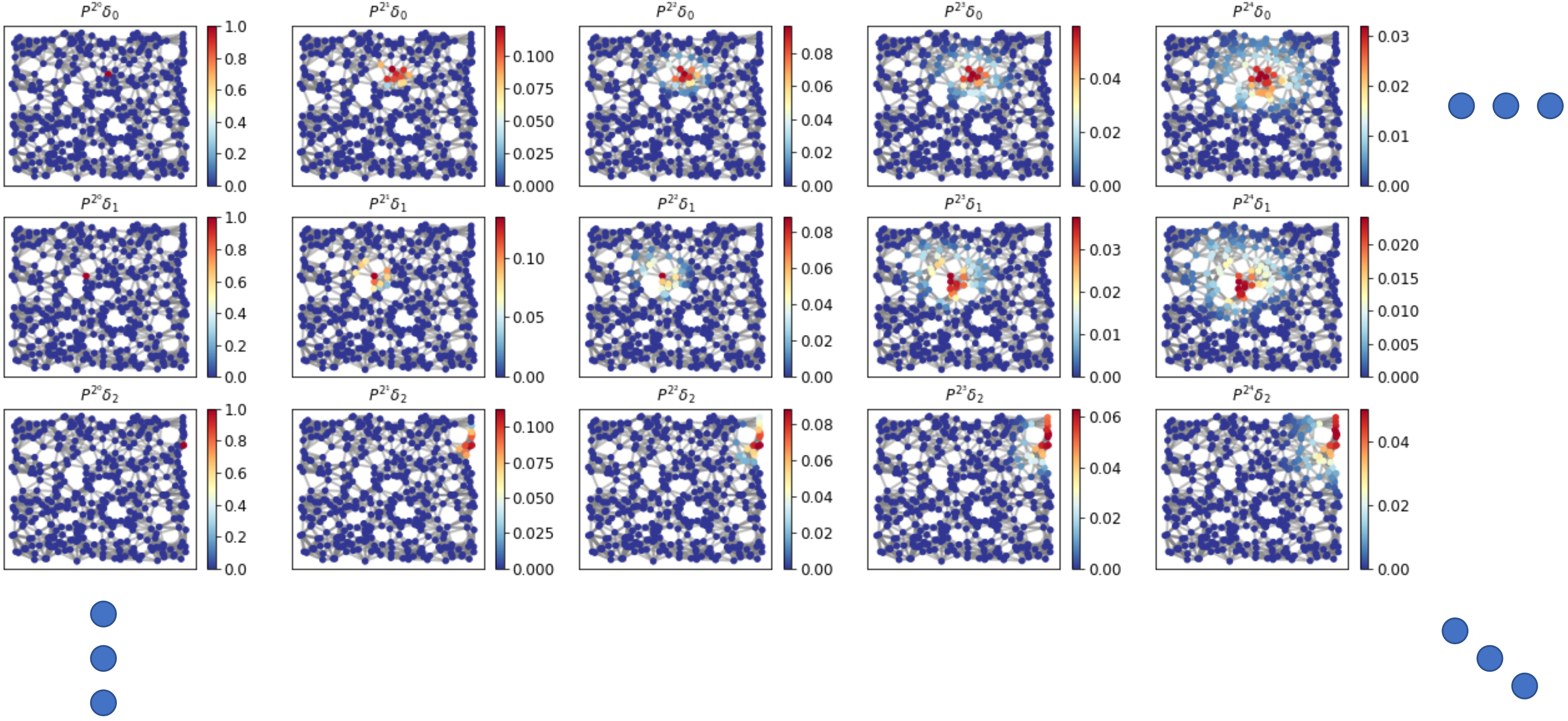


Geometric Scattering
[Gao et al. 2019]

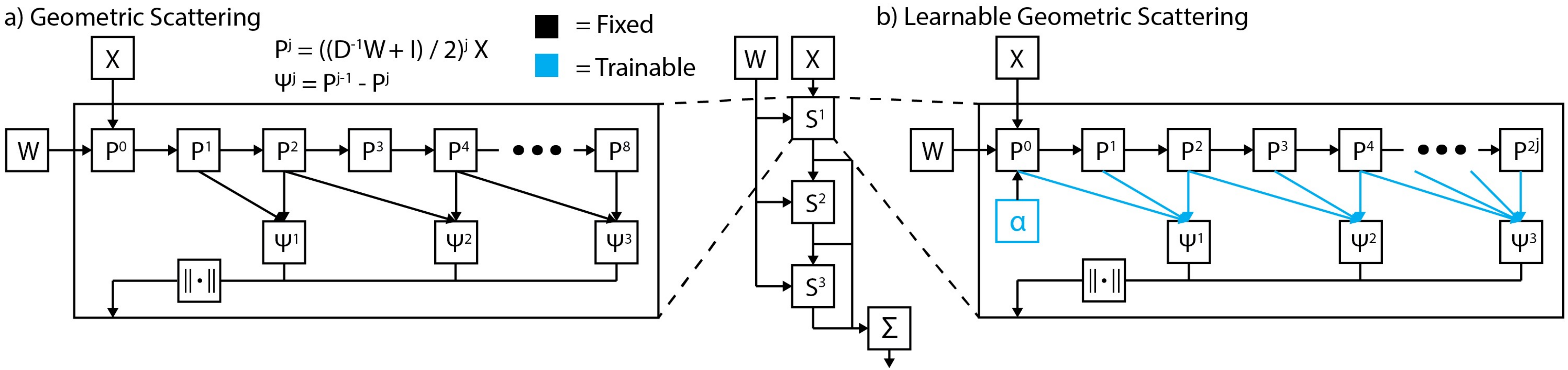
Converts graph signals to vector representation

Provides guarantees on stability and permutation equivariance

Diffusion on a Graphs



Scattering Architecture



$$\boldsymbol{\Psi}_0 := \boldsymbol{I}_n - \boldsymbol{P},$$

$$\boldsymbol{\Psi}_j := \boldsymbol{P}^{2^{j-1}} - \boldsymbol{P}^{2^j} = \boldsymbol{P}^{2^{j-1}}\left(\boldsymbol{I}_n - \boldsymbol{P}^{2^{j-1}}\right), \quad j \geq 1.$$

$$\boldsymbol{U}_p\boldsymbol{x} := \boldsymbol{\Psi}_{j_m}|\boldsymbol{\Psi}_{j_{m-1}} \ldots |\boldsymbol{\Psi}_{j_2}|\boldsymbol{\Psi}_{j_1}\boldsymbol{x}|| \ldots |$$

$$\boldsymbol{S}_{p,q}\boldsymbol{x} := \sum_{i=1}^{n} |\boldsymbol{U}_p\boldsymbol{x}[v_i]|^q$$

Efficient Learning of Diffusion Scales

1. Calculate scales diffusion scales 0 through 16

2. Learn a mixture of scales for each bump using linear layer + Softmax

3. Sort bump from lowest to highest scale by largest weight

4. Calculate wavelets using difference of successive bumps

Results

LEGS-RBF: LEGS followed by radial basis network classifier
LEGS-FCN: LEGS followed by fully connected network
LEGS-FIXED: LEGS with fixed scales followed by fully connected network

	# Graphs	# Classes	Diameter	Nodes	Edges	Clust. Coeff
DD	1178	2	19.81	284.32	715.66	0.48
ENZYMES	600	6	10.92	32.63	62.14	0.45
MUTAG	188	2	8.22	17.93	19.79	0.00
NCI1	4110	2	13.33	29.87	32.30	0.00
NCI109	4127	2	13.14	29.68	32.13	0.00
PROTEINS	1113	2	11.62	39.06	72.82	0.51
PTC	344	2	7.52	14.29	14.69	0.01
COLLAB	5000	3	1.86	74.49	2457.22	0.89
IMDB-BINARY	1000	2	1.86	19.77	96.53	0.95
IMDB-MULTI	1500	3	1.47	13.00	65.94	0.97
REDDIT-BINARY	2000	2	8.59	429.63	497.75	0.05
REDDIT-MULTI-12K	11929	11	9.53	391.41	456.89	0.03
REDDIT-MULTI-5K	4999	5	10.57	508.52	594.87	0.03

	LEGS-RBF	LEGS-FCN	LEGS-FIXED	GCN	GraphSAGE	GS-SVM	Baseline
DD	72.58 ± 3.35	72.07 ± 2.37	69.09 ± 4.82	67.82 ± 3.81	66.37 ± 4.45	72.66 ± 4.94	$\mathbf{75.98 \pm 2.81}$
ENZYMES	36.33 ± 4.50	$\mathbf{38.50 \pm 8.18}$	32.33 ± 5.04	31.33 ± 6.89	15.83 ± 9.10	27.33 ± 5.10	20.50 ± 5.99
MUTAG	33.51 ± 4.34	82.98 ± 9.85	81.84 ± 11.24	79.30 ± 9.66	81.43 ± 11.64	$\mathbf{85.09 \pm 7.44}$	79.80 ± 9.92
NCI1	$\mathbf{74.26 \pm 1.53}$	70.83 ± 2.65	71.24 ± 1.63	60.80 ± 4.26	57.54 ± 3.33	69.68 ± 2.38	56.69 ± 3.07
NCI109	$\mathbf{72.47 \pm 2.11}$	70.17 ± 1.46	69.25 ± 1.75	61.30 ± 2.99	55.15 ± 2.58	68.55 ± 2.06	57.38 ± 2.20
PROTEINS	70.89 ± 3.91	71.06 ± 3.17	67.30 ± 2.94	$\mathbf{74.03 \pm 3.20}$	71.87 ± 3.50	70.98 ± 2.67	73.22 ± 3.76
PTC	$\mathbf{57.26 \pm 5.54}$	56.92 ± 9.36	54.31 ± 6.92	56.34 ± 10.29	55.22 ± 9.13	56.96 ± 7.09	56.71 ± 5.54
COLLAB	75.78 ± 1.95	75.40 ± 1.80	72.94 ± 1.70	73.80 ± 1.73	$\mathbf{76.12 \pm 1.58}$	74.54 ± 2.32	64.76 ± 2.63
IMDB-BINARY	64.90 ± 3.48	64.50 ± 3.50	64.30 ± 3.68	47.40 ± 6.24	46.40 ± 4.03	$\mathbf{66.70 \pm 3.53}$	47.20 ± 5.67
IMDB-MULTI	41.93 ± 3.01	40.13 ± 2.77	41.67 ± 3.19	39.33 ± 3.13	39.73 ± 3.45	$\mathbf{42.13 \pm 2.53}$	39.53 ± 3.63
REDDIT-BINARY	$\mathbf{86.10 \pm 2.92}$	78.15 ± 5.42	85.00 ± 1.93	81.60 ± 2.32	73.40 ± 4.38	85.15 ± 2.78	69.30 ± 5.08
REDDIT-MULTI-12K	38.47 ± 1.07	38.46 ± 1.31	39.74 ± 1.31	$\mathbf{42.57 \pm 0.90}$	32.17 ± 2.04	39.79 ± 1.11	22.07 ± 0.98
REDDIT-MULTI-5K	47.83 ± 2.61	46.97 ± 3.06	47.17 ± 2.93	$\mathbf{52.79 \pm 2.11}$	45.71 ± 2.88	48.79 ± 2.95	36.41 ± 1.80

Results

- Performs well on molecule graphs
  - CASP structure error regression
  - QM9 feature regression

$(\mu \pm \sigma)$	Train MSE	Test MSE
LEGS-FCN	**134.34 ± 8.62**	**144.14 ± 15.48**
LEGS-RBF	140.46 ± 9.76	152.59 ± 14.56
LEGS-FIXED	136.84 ± 15.57	160.03 ± 1.81
GCN	289.33 ± 15.75	303.52 ± 18.90
GraphSAGE	221.14 ± 42.56	219.44 ± 34.84
GIN	221.14 ± 42.56	219.44 ± 34.84
Baseline	393.78 ± 4.02	402.21 ± 21.45

$(\mu \pm \sigma)$	Test MSE
LEGS-FCN	**0.216 ± 0.009**
LEGS-FIXED	0.228 ± 0.019
GraphSAGE	0.524 ± 0.224

$(\mu \pm \sigma)$	Test MSE
GCN	0.417 ± 0.061
GIN	0.247 ± 0.037
Baseline	0.533 ± 0.041

Conclusions

Learnable geometric scattering learns diffusion scales

- More flexible than fixed scattering
- Maintains theoretical properties of fixed scattering
- Improves performance by mitigating oversmoothing and underreaching

Université de Montréal

Mila

Yale

CIFAR

IVADO

Thanks!

Code: https://github.com/KrishnaswamyLab/LearnableScattering

Paper: https://arxiv.org/abs/2010.02415