Scattering Priors for Graph Neural Networks

Alexander Tong[1*], Frederik Wenkel[2*], Kincaid MacDonald[1], Smita Krishnaswamy[1 †], and Guy Wolf[2 †]

[1] Yale University, [2]MILA; University of Montreal, * † Equal contribution

Main Idea

Create a graph neural network based on scattering priors to improve classification accuracy.

Problem Statement

Graph neural networks typically uses cascades of low pass filters.

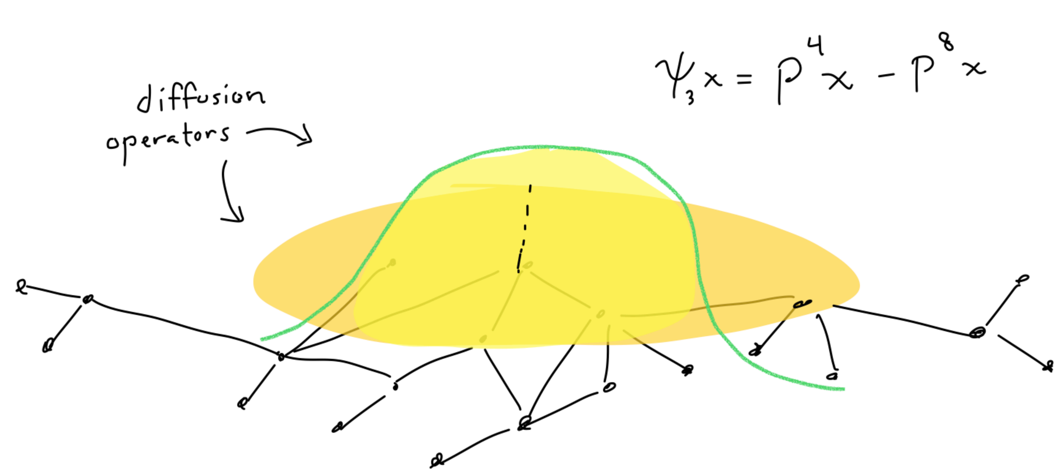
Low pass filters work well in simple data domains but poorly in more complex domains such as are found in biology and chemistry.

Can we improve graph classification with the addition of long range and band pass signals found in scattering transforms?

Background

Graph Diffusion Wavelets [1] Generalizes averaging of points to averaging of distributions based on a ground distance between points.

Allows interpolation of a distribution from a weighted set of distributions.



Geometric Scattering Transform A cascade of diffusion wavelets followed by absolute value non-linearities and averaging operators transforms a graph with features to an embedding space.

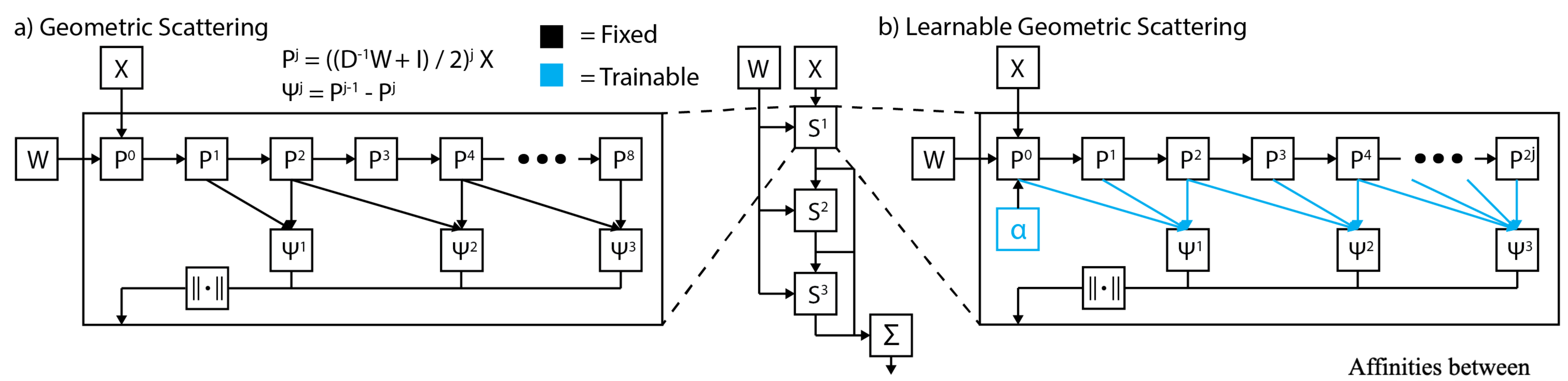
$$\Psi_0 := I_n - P,$$

$$\Psi_j := P^{2^{j-1}} - P^{2^j} = P^{2^{j-1}}(I_n - P^{2^{j-1}}), \quad j \geq 1.$$

$$U_p x := \Psi_{j_m} |\Psi_{j_{m-1}} \dots |\Psi_{j_2} |\Psi_{j_1} x|| \dots |$$

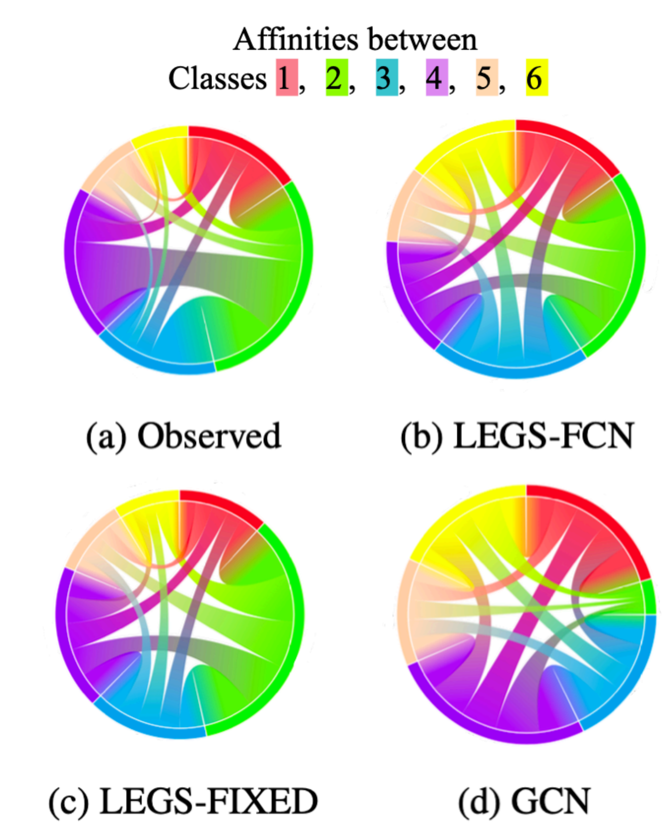
$$S_{p,q} x := \sum_{i=1}^{n} |U_p x[v_i]|^q.$$

Results



	# 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

Statistics of seven biomedical and six social network datasets. Biomedical datasets tend to have a higher ratio of diameter to number of nodes as well as fewer edges for a given size graph.



Enzymes dataset class exchange preferences vs. experimentally observed. LEGS-FCN learns a function which still preserves exchange preferences between classes. Better than the GCN baseline.

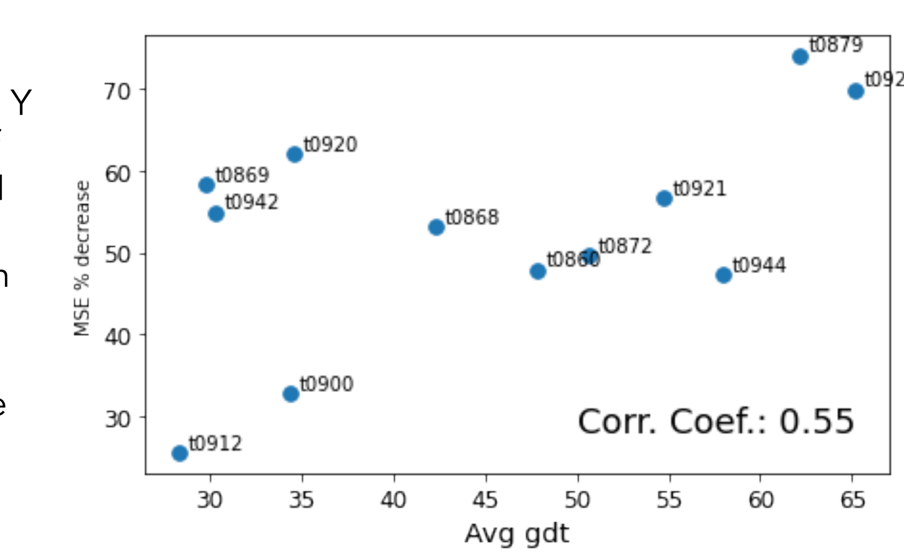
	LEGS-RBF	LEGS-FCN	LEGS-FIXED	GCN	GraphSAGE	GS-SVM	Baseline
DD	72.58 ± 3.35	72.07 ± 2.37	69.09 ± 4.82	67.82 ± 3.81	66.37 ± 4.45	72.66 ± 4.94	$\mathbf{75.98 \pm 2.81}$
ENZYMES	36.33 ± 4.50	$\mathbf{38.50 \pm 8.18}$	32.33 ± 5.04	31.33 ± 6.89	15.83 ± 9.10	27.33 ± 5.10	20.50 ± 5.99
MUTAG	33.51 ± 4.34	82.98 ± 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

Test set accuracy mean +- std. over 10 seeds for 3 LEGS nets and 4 baseline methods.

$(\mu \pm \sigma)$	Train MSE	Test MSE
LEGS-FCN	$\mathbf{134.34 \pm 8.62}$	$\mathbf{144.14 \pm 15.48}$
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
Baseline	393.78 ± 4.02	402.21 ± 21.45

Performance on the critical assessment of structure prediction (CASP) challenge regressing against global distance test (GDT) score. LEGS outperforms on this difficult task by a significant margin on this difficult task.

Performance split by (14) target structures. Y axis is the increase of LEGS-FCN over GCN baseline. LEGS improves the most on targets that are the most complicated as measured by average GDT.



Method

We learn some portions of the geometric scattering transform using standard training methods.

We propose and compare the following models:

LEGS-FIXED – A fixed geometric scattering transform followed by a two layer fully connected network.

LEGS-FCN – A learnable geometric scattering transform followed by a fully connected network.

LEGS-RBF – A learnable geometric scattering transform followed by a custom radial basis network. [1] showed that a radial basis kernel SVM was more effective in classifying from a fixed geometric scattering transform than a linear kernel.

Learnable Geometric Scattering (LEGS net)

(1) the random walk laziness α

(2) The set of scales from dyadic to softmax based

(3) (optionally) add the radial basis network

LEGS Theory

LEGS maintains permutation invariance and robustness to small smooth signal perturbations (shown by non-expansive frame bound) as in the fixed geometric scattering transform as is similarly shown in [3].

Conclusions

Our results show a network based on graph scattering priors outperforms standard GNN architectures particularly on biomedical graph classification datasets.

Our model allows for long range connections with many fewer parameters and maintains useful theoretical properties from fixed geometric scattering transforms.

References

[1] Gao F., Hirn M., Wolf G. Geometric Scattering for Graph Data Analysis. ICML (2019).

[2] Mallat S. Group Invariant Scattering. Communications on Pure and Applied Mathematics (2012).

[3] Perlmutter M., Gao F., Wolf G., Hirn M. Understanding graph neural networks with asymmetric geometric scattering transforms. Arxiv (2019)

Further information

Email: alexander.tong@yale.edu

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

Work supported by IVADO, the Chan-Zuckerberg Initiative, and the NIH.