Geometric Scattering for Graph Data Analysis - Supplement

Feng Gao Guy Wolf Matthew Hirn

A. Detailed graph classification comparison

All results come from the respective papers that introduced the methods, with the exception of: (1) social network results of WL, from Tixier et al. (2017); (2) biochemistry and social results of DCNN, from Verma & Zhang (2018); (3) biochemistry, except for D&D, and social result of GK, from Yanardag & Vishwanathan (2015); (4) D&D of GK is from Niepert et al. (2016); and (5) for Graphlets, biochemistry results from Kriege et al. (2016), social results from Tixier et al. (2017).

B. Detailed tables for scattering feature space analysis from Section 4

Table 2. Classification accuracy with different training/validaion/test splits over scattering features (unnorm. moments)

Dataset	SVM accuracy			
	80%/10%/10%	70%/10%/20%	40%/10%/50%	20%/10%/70%
NCI1	79.80 ± 2.24	78.13 ± 2.07	76.37 ± 0.27	73.60 ± 0.68
NCI109	77.66 ± 1.78	77.54 ± 1.44	74.41 ± 0.14	72.36 ± 0.74
D&D	76.57 ± 3.76	76.74 ± 2.32	76.32 ± 0.59	75.58 ± 0.81
PROTEINS	74.03 ± 4.20	74.30 ± 2.49	73.32 ± 1.68	73.01 ± 1.94
MUTAG	84.04 ± 6.71	82.99 ± 6.97	78.72 ± 3.19	77.47 ± 4.41
PTC	66.32 ± 7.54	64.83 ± 2.13	61.92 ± 1.45	56.75 ± 2.88
ENZYMES	53.83 ± 6.71	52.50 ± 5.35	44.50 ± 3.83	36.38 ± 1.93
COLLAB	76.88 ± 1.13	76.98 ± 0.97	76.42 ± 0.82	74.63 ± 1.05
IMDB-B	70.80 ± 3.54	70.60 ± 2.85	69.10 ± 1.90	67.81 ± 0.98
IMDB-M	48.93 ± 4.77	49.00 ± 1.97	47.20 ± 1.47	44.28 ± 1.87
REDDIT-B	88.30 ± 2.08	88.75 ± 0.96	86.40 ± 0.40	86.18 ± 0.32
REDDIT-5K	50.71 ± 2.27	50.87 ± 1.37	50.10 ± 0.41	48.37 ± 0.76
REDDIT-12K	41.35 ± 1.05	41.05 ± 0.70	39.36 ± 1.30	37.71 ± 0.42

Table 3. Classification accuracy and dimensionality reduction with PCA over scattering features (unnorm. moments)

Dataset	SVM accuracy w.r.t variance covered				PCA dimensions w.r.t variance covered			
	50%	80%	90%	99%	50%	80%	90%	99%
NCI1	72.41 ± 2.36	73.89 ± 2.57	73.89 ± 1.33	78.22 ± 1.95	18	32	43	117
NCI109	70.85 ± 2.59	71.84 ± 2.38	72.33 ± 2.24	76.69 ± 1.02	19	32	43	114
D&D	75.21 ± 3.17	75.13 ± 3.68	74.87 ± 3.99	76.92 ± 3.37	10	35	44	122
PROTEINS	70.80 ± 3.43	74.20 ± 3.06	74.67 ± 3.33	74.57 ± 3.42	2	5	10	36
MUTAG	77.51 ± 10.42	80.32 ± 8.16	82.40 ± 10.92	84.09 ± 9.09	4	8	13	34
PTC	58.17 ± 8.91	60.50 ± 9.96	58.70 ± 6.93	63.68 ± 3.97	7	14	21	62
ENZYMES	29.67 ± 4.46	45.33 ± 6.62	50.67 ± 5.44	52.50 ± 8.89	3	9	16	44
COLLAB	62.86 ± 1.36	71.68 ± 2.06	73.22 ± 2.29	76.54 ± 1.41	2	6	9	32
IMDB-B	58.30 ± 3.44	66.10 ± 3.14	68.80 ± 4.31	68.40 ± 4.31	2	4	8	24
IMDB-M	41.00 ± 4.86	46.40 ± 4.48	45.93 ± 3.86	48.27 ± 3.23	2	5	8	20
REDDIT-B	71.05 ± 2.39	78.95 ± 2.42	83.75 ± 1.83	86.95 ± 1.78	2	5	8	24
REDDIT-5K	40.97 ± 2.06	45.71 ± 2.21	47.43 ± 1.90	49.65 ± 1.86	2	6	10	27
REDDIT-12K	28.22 ± 1.64	33.36 ± 0.93	34.71 ± 1.52	38.39 ± 1.54	2	5	9	27

Table 1. Comparison of the proposed graph scattering classifier (GSC) with graph kernel methods and deep learning methods on biochemistry & social graph datasets. (Remark[1]: DCNN using different training/test split)

	Graph kernel			Deep learning				
	NCI1	NCI109	D&D	PROTEINS	MUTAG	PTC	ENZYMES
WL	84.46 ± 0.45	85.12 ± 0.29	78.34 ± 0.62	72.92 ± 0.56	84.11 ± 1.91	59.97 ± 1.60	55.22 ± 1.26
PK	82.54 ± 0.47	N/A	78.25 ± 0.51	73.68 ± 0.68	76.00 ± 2.69	59.50 ± 2.44	N/A
Graphlet	70.5 ± 0.2	69.3 ± 0.2	79.7 ± 0.7	72.7 ± 0.6	85.2 ± 0.9	54.7 ± 2.0	30.6 ± 1.2
WL-OA	86.1 ± 0.2	86.3 ± 0.2	79.2 ± 0.4	76.4 ± 0.4	84.5 ± 1.7	63.6 ± 1.5	59.9 ± 1.1
GK	62.28 ± 0.29	62.60 ± 0.19	78.45 ± 0.26	71.67 ± 0.55	81.39 ± 1.74	57.26 ± 1.41	26.61 ± 0.99
DGK	80.3 ± 0.4	80.3 ± 0.3	73.09 ± 0.25	75.7 ± 0.50	87.4 ± 2.7	60.1 ± 2.5	53.4 ± 0.9
DGCNN	74.44 ± 0.47	N/A	79.37 ± 0.94	75.54 ± 0.94	85.83 ± 1.66	58.59 ± 2.47	51.00 ± 7.29
graph2vec	73.22 ± 1.81	74.26 ± 1.47	N/A	73.30 ± 2.05	83.15 ± 9.25	60.17 ± 6.86	N/A
2D CNN	N/A	N/A	N/A	77.12 ± 2.79	N/A	N/A	N/A
CCN	76.27 ± 4.13	75.54 ± 3.36	N/A	N/A	91.64 ± 7.24	70.62 ± 7.04	N/A
PSCN ($k = 10$)	76.34 ± 1.68	N/A	76.27 ± 2.15	75.00 ± 2.51	88.95 ± 4.37	62.29 ± 5.68	N/A
DCNN	56.61 ± 1.04	57.47 ± 1.22	58.09 ± 0.53	61.29 ± 1.60	56.60 ± 2.89	56[1]	42.44 ± 1.76
GCAPS-CNN	82.72 ± 2.38	81.12 ± 1.28	77.62 ± 4.99	76.40 ± 4.17	N/A	66.01 ± 5.91	61.83 ± 5.39
S2S-P2P-NN	83.72 ± 0.4	83.64 ± 0.3	N/A	76.61 ± 0.5	89.86 ± 1.1	64.54 ± 1.1	63.96 ± 0.6
GIN-0 (MLP-SUM)	82.70 ± 1.60	N/A	N/A	76.20 ± 2.80	89.40 ± 5.60	64.60 ± 7.00	N/A
GS-SVM	79.14 ± 1.28	77.95 ± 1.25	75.04 ± 3.64	74.11 ± 4.02	83.57 ± 6.75	63.94 ± 7.38	56.83 ± 4.97

	Graph kernel			Deep learning			
	COLLAB	IMDB-B	IMDB-M	REDDIT-B	REDDIT-5K	REDDIT-12K
WL	77.82 ± 1.45	71.60 ± 5.16	N/A	78.52 ± 2.01	50.77 ± 2.02	34.57 ± 1.32
PK	N/A	N/A	N/A	N/A	N/A	N/A
Graphlet	73.42 ± 2.43	65.4 ± 5.95	N/A	77.26 ± 2.34	39.75 ± 1.36	25.98 ± 1.29
WL-OA	80.7 ± 0.1	N/A	N/A	89.3 ± 0.3	N/A	N/A
GK	72.84 ± 0.28	65.87 ± 0.98	43.89 ± 0.38	77.34 ± 0.18	41.01 ± 0.17	N/A
DGK	73.0 ± 0.2	66.9 ± 0.5	44.5 ± 0.5	78.0 ± 0.3	41.2 ± 0.1	32.2 ± 0.1
DGCNN	73.76 ± 0.49	70.03 ± 0.86	47.83 ± 0.85	N/A	48.70 ± 4.54	N/A
graph2vec	N/A	N/A	N/A	N/A	N/A	N/A
2D CNN	71.33 ± 1.96	70.40 ± 3.85	N/A	89.12 ± 1.7	52.21 ± 2.44	48.13 ± 1.47
CCN	N/A	N/A	N/A	N/A	N/A	N/A
PSCN ($k = 10$)	72.60 ± 2.15	71.00 ± 2.29	45.23 ± 2.84	86.30 ± 1.58	49.10 ± 0.7	41.32 ± 0.42
DCNN	52.11 ± 0.71	49.06 ± 1.37	33.49 ± 1.42	N/A	N/A	N/A
GCAPS-CNN	77.71 ± 2.51	71.69 ± 3.40	48.50 ± 4.1	87.61 ± 2.51	50.10 ± 1.72	N/A
S2S-P2P-NN	81.75 ± 0.8	73.8 ± 0.7	51.19 ± 0.5	86.50 ± 0.8	52.28 ± 0.5	42.47 ± 0.1
GIN-0 (MLP-SUM)	80.20 ± 1.90	75.10 ± 5.10	52.30 ± 2.80	92.40 ± 2.50	57.50 ± 1.50	N/A
GS-SVM	79.94 ± 1.61	71.20 ± 3.25	48.73 ± 2.32	89.65 ± 1.94	53.33 ± 1.37	45.23 ± 1.25

Table 4. Dimensionality reduction with PCA over scattering features (unnorm. moments)

Dataset	SVM accuracy		PCA dimensions ($> 90\%$ variance)						
	PCA	Full	All classes	Per class					
ENZYMES	50.67 ± 5.44	53.83 ± 6.71	16	9	8	8	9	10	6

Table 5. EC subspace analysis in scattering feature space of ENZYMES (Borgwardt et al., 2005)

Enzyme Class:	Mean distance to subspace of class						True class as		
	EC-1	EC-2	EC-3	EC-4	EC-5	EC-6	1^{st}	2^{nd}	3^{rd}-6^{th}
	measured via PCA projection/reconstruction distance						nearest subspace		
EC-1	18.15	98.44	75.47	62.87	53.07	84.86	45%	28%	27%
EC-2	22.65	9.43	30.14	22.66	18.45	22.75	53%	24%	23%
EC-3	107.23	252.31	30.4	144.08	117.24	168.56	32%	7%	61%
EC-4	117.68	127.27	122.3	29.59	94.3	49.14	24%	12%	64%
EC-5	45.46	66.57	60	50.07	15.09	58.22	67%	21%	12%
EC-6	62.38	58.88	73.96	51.94	59.23	13.56	67%	21%	12%

C. Detailed Dataset Descriptions

The details of the datasets used in this work are as follows:

NCI1 (Wale et al., 2008) contains 4,110 chemical compounds as graphs, with 37 node features. Each compound is labeled according to is activity against non-small cell lung cancer and ovarian cancer cell lines, and these labels serve as classification goal on this data.

NCI109 (Wale et al., 2008) is similar to NCI1, but with 4,127 chemical compounds and 38 node features.

MUTAG (Debnath et al., 1991) consists of 188 mutagenic aromatic and heteroaromatic nitro compounds (as graphs) with 7 node features. The classification here is binary (i.e., two classes), based on whether or not a compound has a mutagenic effect on bacterium.

PTC (Toivonen et al., 2003) is a dataset of 344 chemical compounds (as graphs) with nineteen node features that are divided into two classes depending on whether they are carcinogenic in rats.

PROTEINS (Borgwardt et al., 2005) dataset contains 1,113 proteins (as graphs) with three node features, where the goal of the classification is to predict whether the protein is enzyme or not.

D&D (Dobson & Doig, 2003) dataset contains 1,178 protein structures (as graphs) that, similar to the previous one, are classified as enzymes or non-enzymes.

ENZYMES (Borgwardt et al., 2005) is a dataset of 600 protein structures (as graphs) with three node features. These proteins are divided into six classes of enzymes (labelled by enzyme commission numbers) for classification.

COLLAB (Yanardag & Vishwanathan, 2015) is a scientific collaboration dataset contains 5K graphs. The classification goal here is to predict whether the graph belongs to a subfield of Physics.

IMDB-B (Yanardag & Vishwanathan, 2015) is a movie collaboration dataset with contains 1K graphs. The graphs are generated on two genres: Action and Romance, the classification goal is to predict the correct genre for each graph.

IMDB-M (Yanardag & Vishwanathan, 2015) is similar to IMDB-B, but with 1.5K graphs & 3 genres: Comedy, Romance, and Sci-Fi.

REDDIT-B (Yanardag & Vishwanathan, 2015) is a dataset with 2K graphs, where each graph corresponds to an online discussion thread. The classification goal is to predict whether the graph belongs to a Q&A-based community or discussion-based community.

REDDIT-5K (Yanardag & Vishwanathan, 2015) consists of 5K threads (as graphs) from five different subreddits. The classification goal is to predict the corresponding subreddit for each thread.

REDDIT-12K (Yanardag & Vishwanathan, 2015) is similar to REDDIT-5k, but with 11,929 graphs from 12 different subreddits.

Table 6 summarizes the size of available graph data (i.e., number of graphs, and both max & mean number of vertices within graphs) in these datasets, as previously reported in the literature.

Table 6. Basic statistics of the graph classification databases

	NCI1	NCI109	MUTAG	D&D	PTC	PROTEINS
# of graphs in data:	4110	4127	188	1178	344	1113
Max # of vertices:	111	111	28	5748	109	620
Mean # of vertices:	29.8	29.6	17.93	284.32	25.56	39.0
# of features per vertex:	37	38	7	89	22	3
Mean # of edges:	64.6	62.2	39.50	1431.3	51.90	72.82
# of classes:	2	2	2	2	2	2

ENZYMES	COLLAB	IMDB		REDDIT		
		B	M	B	5K	12K
600	5000	1000	1500	2000	5000	11929
126	492	136	89	3783	3783	3782
32.6	74.49	19.77	13	429.61	508.5	391.4
3	3	3	3	2	2	2
124.2	2457.78	96.53	65.94	497.75	594.87	456.89
6	3	2	3	2	5	11

Graph signals for social network data: None of the social network datasets has ready-to-use node features. Therefore, in the case of COLLAB, IMDB-B, and IMDB-M, we use the eccentricity and clustering coefficients for each vertex as characteristic graph signals. In the case of REDDIT-B, REDDIT-5K and REDDIT-12K, on the other hand, we only use the clustering coefficient, due to the presence of disconnected graphs in these datasets.

D. Technical Details

The computation of the scattering features is based on several design choices, akin to typical architecture choices in neural networks. Most importantly, it requires a choice of 1. which statistical moments to use (normalized or unnormalized), 2. the number of wavelet scales to use (given by J), and 3. the number of moments to use (denoted by Q). In general, J can be automatically tuned by the diameter of the considered graphs (e.g., setting it to the logarithm of the diameter), and the other choices can be tuned via cross-validation. However, we have found the impact of such tuning to be minor, and thus for simplicity, we fix our configuration to use normalized moments, $J = 5$, and $Q = 4$ throughout this work.

Cross validation procedure: Classification evaluation was done with standard ten-fold cross validation procedure. First, the entire dataset is randomly split into ten subsets. Then, in each iteration (or "fold"), nine of them are used as training and validation, and the other one is used for testing classification accuracy. In total, after ten iterations, each of the subsets has been used once for testing, resulting in ten reported classification accuracy numbers for the examined dataset. Finally, the mean and standard deviation of these ten accuracies are computed and reported.

It should be noted that during training, each iteration also performs automatic tuning of the trained classifier, as follows. First, nine iterations are performed, each time using eight subsets (i.e., folds) as training and the remaining one as validation set, which is used to determine the optimal parameters for SVM. After nine iterations, each of the training/validation subsets has been used once for validation, and we obtain nine classification models, which in turn produce nine predictions (i.e., class assignments) for each data point in the test subset of the main cross validation. To obtain the final predicted class of this cross validation iteration, we select the class with the most votes (from among the nine models) as our final classification result. These results are then compared to the true labels (in the test set) on the test subset to obtain classification accuracy for this fold.

Software & hardware environment: Geometric scattering and related classification code were implemented in Python. All experiments were performed on HPC environment using an intel16-k80 cluster, with a job requesting one node with four

processors and two Nvidia Tesla k80 GPUs.

E. Ablation Study

To fully understand the power of our geometric scattering coefficients, we conduct an ablation study using five social network datasets, namely COLLAB, IMDB-B, IMDB-M, REDDIT-B, REDDIT-5K, as representative examples. Following the settings in the main paper, here instead of using four normalized moments for each order of scattering moments, we only use one normalized moment (mean) and two normalized moments (mean and variance) and compare the graph classification results in Table 7. We show that using only one normalized moment our method can still get relatively good results, and using higher order moments helps us to match or outperform most state-of-the-art results. Generally, the results degrade by 1-6% on the social network data sets reducing from using four normalized moments to two or one normalized moment.

Table 7. Ablation study on five social network datasets using only one normalized moments and two normalized moments.

	COLLAB	IMDB-B	IMDB-M	REDDIT-B	REDDIT-5K
One normalized moment	77.42	69.80	48.47	83.25	50.31
Two normalized moments	78.44	69.3	48.27	85.20	51.49

Finally, we perform graph classification with two different classifiers: linear SVM and fully connected layers (FCLs)[1] to further demonstrate the usefulness of geometric scattering coefficients and show that our scattering coefficients perform well regardless of the choice of classifiers. Our results in Table 8 show that compared to RBF SVM, FCLs and linear SVM are worse (1-3%) but not by too much.

Table 8. Graph classfication with FCLs and linear SVM classifiers

	COLLAB	IMDB-B	IMDB-M	REDDIT-B	REDDIT-5K
linear SVM	77.40	70.50	47.13	86.45	53.23
FCLs	79.26	69.50	46.40	86.60	50.50

References

Borgwardt, K. M., Ong, C. S., Schönauer, S., Vishwanathan, S., Smola, A. J., and Kriegel, H.-P. Protein function prediction via graph kernels. *Bioinformatics*, 21(suppl_1):i47–i56, 2005.

Debnath, A. K., Lopez de Compadre, R. L., Debnath, G., Shusterman, A. J., and Hansch, C. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity. *Journal of medicinal chemistry*, 34(2):786–797, 1991.

Dobson, P. D. and Doig, A. J. Distinguishing enzyme structures from non-enzymes without alignments. *Journal of molecular biology*, 330(4):771–783, 2003.

Kriege, N. M., Giscard, P.-L., and Wilson, R. On valid optimal assignment kernels and applications to graph classification. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 29*, pp. 1623–1631. Curran Associates, Inc., 2016.

Niepert, M., Ahmed, M., and Kutzkov, K. Learning convolutional neural networks for graphs. In *International conference on machine learning*, pp. 2014–2023, 2016.

Tixier, A. J.-P., Nikolentzos, G., Meladianos, P., and Vazirgiannis, M. Classifying graphs as images with convolutional neural networks. arXiv:1708.02218, 2017.

Toivonen, H., Srinivasan, A., King, R. D., Kramer, S., and Helma, C. Statistical evaluation of the predictive toxicology challenge 2000–2001. *Bioinformatics*, 19(10):1183–1193, 2003.

Verma, S. and Zhang, Z.-L. Graph capsule convolutional neural networks. In *Joint ICML and IJCAI Workshop on Computational Biology*, 2018. arXiv:1805.08090.

[1]Hyperparameters of FCLs are manually selected

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.

Yanardag, P. and Vishwanathan, S. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1365–1374. ACM, 2015.