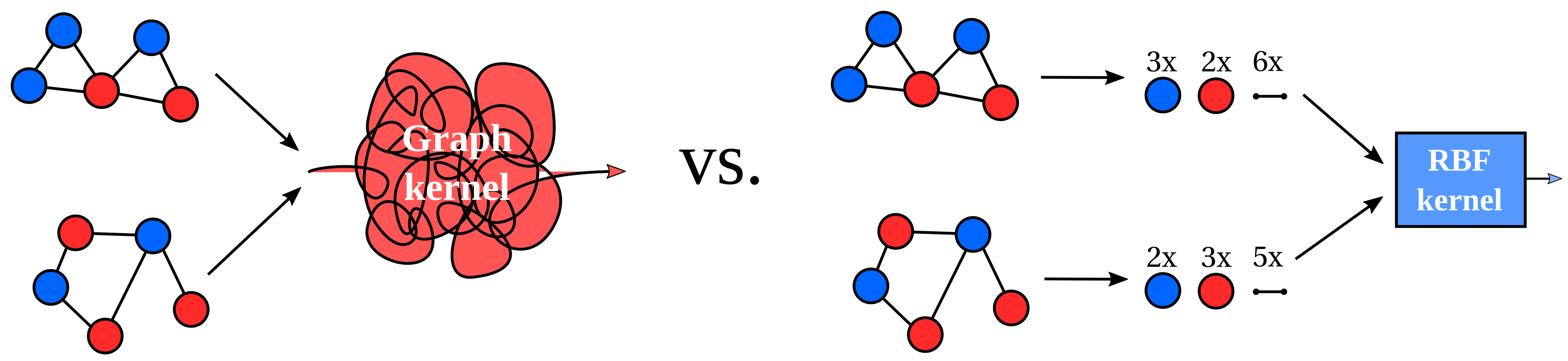
On the Necessity of Graph Kernel Baselines

Till Hendrik Schulz, Pascal Welke

UNIVERSITÄT BONN

Is graph structure even relevant in classification tasks of benchmark datasets?

- We compare established graph kernels to a **kernel which disregards *all* graph structure**.
- The **No-Graph kernel (NoG)** considers graphs as a multiset of vertex and edge labels.



How does NoG perform compared to more sophisticated graph kernels?

Tested kernels:

NoG - No-Graph baseline kernel
psf - Probabilistic frequent subtree kernel [5]
bpsf - Boosted probabilistic frequent subtree kernel [5]
fsg - Frequent subgraph kernel based on FSG [7]
cp - Cyclic pattern kernel [2]
gs - Graphlet sampling kernel [4]
sp - Shortest path kernel [1]
rw - Random walk kernel [6]
wl - Weisfeiler Lehman kernel [3]

Evaluation details:

The predictive performance was measured in terms of accuracy obtained by SVMs using a 10-fold cross-validation. The kernel and SVM parameters were identified using an extensive grid search.

Legend:

- no significant difference to NoG
- kernel performs significantly worse than NoG
- kernel performs significantly better than NoG

x - result unavailable due to time/memory constraints

Observations & Interpretations:

- No tested graph kernel achieves results significantly better than the baseline on more than a few datasets.
- Graph kernels can hardly prove their functionality on available datasets.
- Utilizing the graphs' structure in graph kernels does not significantly improve the classification accuracy.
- Graph structure may not even be relevant to perform well on a majority of benchmark datasets.

	NoG	psf	bpsf	fsg	cp	gs	sp	rw	wl
AIDS	99.65	98.25	98.45	97.85	98.60				98.70
BZR	86.16				78.78	78.53		73.82	
BZR_MD	70.19		58.14	x	x	52.34		50.66	60.75
COIL-DEL	14.13	7.83	7.68	x	x	9.74		x	
COIL-RAG	7.22	3.38	3.36	5.69	x	3.16	6.61	x	
COX2	81.37	78.16	78.16	69.39	77.95	78.16		77.95	
COX2_MD	65.26				x	51.81		51.15	
DD	76.67				x		x	x	
DHFR	74.06				54.77	67.05		60.98	82.02
DHFR_MD	64.88			x	x		69.20		
ENZYMES	43.33	28.33	32.00	x	x	30.50		17.33	50.67
ER_MD	74.46				x	59.42	59.21	59.42	
IMDB-BINARY	70.70	59.00	60.10	61.50	x	63.90	47.90	x	
IMDB-MULTI	46.73	38.93	40.13		x	39.53	34.13	x	
Letter-high	34.58	29.42	28.89		x	18.70	30.47		
Letter-low	43.02	27.29	27.07	27.29	x	14.93	48.60	47.04	39.91
Letter-med	38.71	26.80	27.07	26.80	x	14.51	44.67		
MSRC_21	86.46	51.84	52.22	46.79	x	14.60		5.06	
MSRC_21C	81.59			64.63	x	16.43		6.43	
MSRC_9	88.35			x	x	25.09		11.73	
MUTAG	87.31								
Mutagenicity	75.56				79.06	64.61		x	83.56
NCI1	69.93		74.33	76.28		62.68		x	84.72
NCI109	68.48	73.56	72.64	75.67	73.56	64.67		x	85.20
PROTEINS	74.58			x	x			x	
PROTEINS_full	74.58			x	x			x	
PTC_FM	63.63	58.17				57.08	57.85		
PTC_FR	67.25					65.53		65.25	
PTC_MM	66.70				61.93	61.31		61.62	
PTC_MR	57.60								
REDDIT-BINARY	83.50	55.75	57.70	x	x	x	x	x	72.80
REDDIT-MULTI-12K	36.99	23.18	23.12	x	x	x	x	x	x
REDDIT-MULTI-5K	49.81	22.46	22.90	x	x	x	x	x	x
SYNTHETIC	50.00			x	x				
SYNTHETICnew	64.33	51.33			x			54.00	99.33
Synthie	50.69		42.28	x	x	41.32		16.05	
Tox21_AHR	90.89	88.47	88.36	88.37	x	88.38	88.41	x	93.36
Tox21_AR-LBD	98.05				x	96.50	97.50	x	98.49
Tox21_ARE	86.42	84.79	84.74	84.69	x	84.67	84.75	x	88.94

What do the results suggest?

- Most available datasets aren't suitable for benchmarking purposes.
- New benchmark datasets that highlight the power of graph kernels are necessary.
- Graph kernel baselines are imperative in order to put graph kernel performances into context.

References

[1] K. M. Borgwardt and H.-P. Kriegel. Shortest-path kernels on graphs. In *IEEE International Conference on Data Mining (ICDM) Proceedings*, pages 74–81. IEEE Computer Society, 2005.

[2] T. Horváth, T. Gärtner, and S. Wrobel. Cyclic pattern kernels for predictive graph mining. In W. Kim, R. Kohavi, J. Gehrke, and W. DuMouchel, editors, *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Proceedings*, pages 158–167, 2004.

[3] N. Shervashidze, P. Schweitzer, E. J. Van Leeuwen, K. Mehlhorn, and K. M. Borgwardt. Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research*, 12:2539–2561, 2011.

[4] N. Shervashidze, S. V. N. Vishwanathan, T. Petri, K. Mehlhorn, and K. M. Borgwardt. Efficient graphlet kernels for large graph comparison. In D. A. V. Dyk and M. Welling, editors, *International Conference on Artificial Intelligence and Statistics (AISTATS) Proceedings*, pages 488–495, 2009.

[5] P. Welke, T. Horváth, and S. Wrobel. Probabilistic frequent subtrees for efficient graph classification and retrieval. *Machine Learning*, 107(11):1847–1873, 2018.

[6] T. Gärtner, P. Flach, and S. Wrobel. On graph kernels: Hardness results and efficient alternatives. In B. Schölkopf and M. K. Warmuth, editors, *Learning Theory and Kernel Machines*, pages 129–143. Springer, 2003.

[7] M. Kuramochi and G. Karypis. An efficient algorithm for discovering frequent subgraphs. *Transactions on Knowledge and Data Engineering*, 16(9):1038–1051, 2004.