

Table 1. Summary of graph data set characteristics, viz. the size of the training (tr), the validation (va) and the test set (te), the number of classes ($\#classes$), the label alphabet of both nodes and edges, the average and maximum number of nodes and edges (\emptyset/\max nodes/edges), whether the graphs are uniformly distributed over the classes or not (balanced), and the recognition rate of the k -NN classifier (RR).

Database	size (tr, va, te)	$\#classes$	node labels	edge labels	\emptyset nodes	\emptyset edges	max nodes	max edges	balanced	RR
Letter (<i>low</i>)	750, 750, 750	15	x, y coordinates	none	4.7	3.1	8	6	Y	99.6%
Letter (<i>medium</i>)	750, 750, 750	15	x, y coordinates	none	4.7	3.2	9	7	Y	94.0%
Letter (<i>high</i>)	750, 750, 750	15	x, y coordinates	none	4.7	4.5	9	9	Y	90.0%
Digit	1,000, 500, 2,000	10	x, y coordinates	Angle	11.8	13.1	32	30	Y	91.0%
GREC	286, 286, 528	22	x, y coordinates	Line type	11.5	12.2	25	30	Y	95.5%
Fingerprint	500, 300, 2,000	4	x, y coordinates	Angle	5.42	4.42	26	24	N	76.6%
COIL-RAG	2,400, 500, 1,000	100	RGB histogram	Boundary length	3.0	3.0	11	13	Y	92.5%
COIL-DEL	2,400, 500, 1,000	100	x, y coordinates	none	21.5	54.2	77	222	Y	93.3%
Web	780, 780, 780	20	Word and its frequency	Section(s) type	186.1	104.6	834	596	N	80.3%
AIDS	250, 250, 1,500	2	Chemical symbol	Valence	15.7	16.2	95	103	N	97.3%
Mutagenicity	1,500, 500, 2,337	2	Chemical symbol	Valence	30.3	30.8	417	112	N	71.5%
Protein	200, 200, 200	6	Type and aa-sequence	Type and distance	32.6	62.1	126	149	Y	65.5%