

Dataset	5% CS						KMEANS	GT
	PCCC	PCCC-R	COPKM	CSC	DILS	LCC		
Appendicitis	<b>0</b>	<b>0</b>	<b>0</b>	1	<b>0</b>	<b>0</b>	1	0
Breast Cancer	<b>0</b>	<b>0</b>	–	189	3	23	39	0
Bupa	<b>0</b>	<b>0</b>	–	70	1	6	58	0
Circles	<b>0</b>	<b>0</b>	–	49	<b>0</b>	6	31	0
Ecoli	<b>0</b>	<b>0</b>	<b>0</b>	70	<b>0</b>	<b>0</b>	21	0
Glass	<b>0</b>	<b>0</b>	<b>0</b>	36	<b>0</b>	<b>0</b>	11	0
Haberman	<b>0</b>	<b>0</b>	–	42	<b>0</b>	4	28	0
Hayesroth	<b>0</b>	<b>0</b>	<b>0</b>	14	<b>0</b>	1	6	0
Heart	<b>0</b>	<b>0</b>	–	49	<b>0</b>	1	18	0
Ionosphere	<b>0</b>	<b>0</b>	–	61	1	8	23	0
Iris	<b>0</b>	<b>0</b>	<b>0</b>	15	<b>0</b>	<b>0</b>	2	0
Led7Digit	<b>0</b>	<b>0</b>	<b>0</b>	210	<b>0</b>	–	18	0
Monk2	<b>0</b>	<b>0</b>	–	123	1	15	64	0
Moons	<b>0</b>	<b>0</b>	<b>0</b>	44	<b>0</b>	1	15	0
Movement Libras	<b>0</b>	<b>0</b>	<b>0</b>	100	<b>0</b>	<b>0</b>	6	0
Newthyroid	<b>0</b>	<b>0</b>	<b>0</b>	29	<b>0</b>	<b>0</b>	11	0
Saheart	<b>0</b>	<b>0</b>	–	122	6	25	55	0
Sonar	<b>0</b>	<b>0</b>	–	25	<b>0</b>	1	14	0
Soybean	<b>0</b>	<b>0</b>	<b>0</b>	1	<b>0</b>	<b>0</b>	0	0
Spectfheart	<b>0</b>	<b>0</b>	–	32	<b>0</b>	2	29	0
Spiral	<b>0</b>	<b>0</b>	–	51	<b>0</b>	5	29	0
Tae	<b>0</b>	<b>0</b>	<b>0</b>	15	<b>0</b>	<b>0</b>	7	0
Vehicle	<b>0</b>	<b>0</b>	–	603	7	87	204	0
Wine	<b>0</b>	<b>0</b>	<b>0</b>	25	<b>0</b>	<b>0</b>	2	0
Zoo	<b>0</b>	<b>0</b>	<b>0</b>	2	<b>0</b>	<b>0</b>	0	0
Mean	<b>0</b>	<b>0</b>	<b>0*</b>	79	1	8*	28	0

\*Nan values (–) are ignored when computing the sum.

Table W17: Average number of cannot-link constraint violations of the PCCC and the PCCC-R algorithms and the four state-of-the-art algorithms (COPKM, CSC, DILS, LCC) for the constraint sets of size 5% CS. The lowest values are stated in bold. The column KMEANS reports the average number of cannot-link constraint violations obtained with the k-means algorithm. The hyphen indicates that the respective algorithm returned no solution within the time limit of 1,800 seconds.