

Dataset	10% CS						KMEANS	GT
	PCCC	PCCC-R	COPKM	CSC	DILS	LCC		
Appendicitis	<b>0</b>	<b>0</b>	–	11	<b>0</b>	5	5	0
Breast Cancer	<b>0</b>	<b>0</b>	–	717	<b>0</b>	85	133	0
Bupa	<b>0</b>	<b>0</b>	–	261	<b>0</b>	44	211	0
Circles	<b>0</b>	<b>0</b>	–	<b>0</b>	<b>0</b>	39	109	0
Ecoli	<b>0</b>	<b>0</b>	<b>0</b>	74	1	3	34	0
Glass	<b>0</b>	<b>0</b>	<b>0</b>	118	<b>0</b>	13	57	0
Haberman	<b>0</b>	<b>0</b>	<b>0</b>	159	<b>0</b>	21	92	0
Hayesroth	<b>0</b>	<b>0</b>	–	69	<b>0</b>	11	31	0
Heart	<b>0</b>	<b>0</b>	–	156	<b>0</b>	75	44	0
Ionosphere	<b>0</b>	<b>0</b>	–	245	<b>0</b>	93	126	0
Iris	<b>0</b>	<b>0</b>	<b>0</b>	21	<b>0</b>	<b>0</b>	1	0
Led7Digit	<b>0</b>	<b>0</b>	<b>0</b>	512	4	24	85	0
Monk2	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	99	239	0
Moons	<b>0</b>	<b>0</b>	–	<b>0</b>	<b>0</b>	19	78	0
Movement Libras	<b>0</b>	<b>0</b>	<b>0</b>	68	<b>0</b>	7	24	0
Newthyroid	<b>0</b>	<b>0</b>	<b>0</b>	69	1	21	33	0
Saheart	<b>0</b>	<b>0</b>	<b>0</b>	52	<b>0</b>	118	226	0
Sonar	<b>0</b>	<b>0</b>	–	108	<b>0</b>	32	65	0
Soybean	<b>0</b>	<b>0</b>	<b>0</b>	5	<b>0</b>	<b>0</b>	0	0
Spectfheart	<b>0</b>	<b>0</b>	–	113	<b>0</b>	31	93	0
Spiral	<b>0</b>	<b>0</b>	–	84	1	45	95	0
Tae	<b>0</b>	<b>0</b>	–	51	<b>0</b>	5	24	0
Vehicle	<b>0</b>	<b>0</b>	–	303	164	–	760	0
Wine	<b>0</b>	<b>0</b>	<b>0</b>	24	<b>0</b>	1	8	0
Zoo	<b>0</b>	<b>0</b>	<b>0</b>	13	<b>0</b>	<b>0</b>	0	0
Mean	<b>0</b>	<b>0</b>	<b>0*</b>	129	7	33*	103	0

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

Table W14: 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 10% 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.