

Dataset	15% CS						KMEANS	GT
	PCCC	PCCC-R	COPKM	CSC	DILS	LCC		
Appendicitis	<b>0</b>	<b>0</b>	–	1	1	18	22	0
Breast Cancer	<b>0</b>	<b>0</b>	<b>0</b>	1,676	<b>0</b>	<b>0</b>	337	0
Bupa	<b>0</b>	<b>0</b>	<b>0</b>	608	<b>0</b>	–	464	0
Circles	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	253	0
Ecoli	<b>0</b>	<b>0</b>	–	92	7	38	114	0
Glass	<b>0</b>	<b>0</b>	–	272	1	36	125	0
Haberman	<b>0</b>	<b>0</b>	<b>0</b>	394	<b>0</b>	12	252	0
Hayesroth	<b>0</b>	<b>0</b>	–	67	<b>0</b>	31	54	0
Heart	<b>0</b>	<b>0</b>	<b>0</b>	141	<b>0</b>	39	145	0
Ionosphere	<b>0</b>	<b>0</b>	<b>0</b>	554	<b>0</b>	318	272	0
Iris	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	1	0
Led7Digit	<b>0</b>	<b>0</b>	–	895	27	–	178	0
Monk2	<b>0</b>	<b>0</b>	<b>0</b>	13	<b>0</b>	<b>0</b>	539	0
Moons	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	40	128	0
Movement Libras	<b>0</b>	<b>0</b>	<b>0</b>	128	3	34	77	0
Newthyroid	<b>0</b>	<b>0</b>	<b>0</b>	76	2	49	87	0
Saheart	<b>0</b>	<b>0</b>	<b>0</b>	123	<b>0</b>	–	495	0
Sonar	<b>0</b>	<b>0</b>	–	247	<b>0</b>	27	135	0
Soybean	<b>0</b>	<b>0</b>	<b>0</b>	13	<b>0</b>	<b>0</b>	1	0
Spectfheart	<b>0</b>	<b>0</b>	<b>0</b>	263	<b>0</b>	<b>0</b>	225	0
Spiral	<b>0</b>	<b>0</b>	–	2	<b>0</b>	21	239	0
Tae	<b>0</b>	<b>0</b>	–	49	2	40	62	0
Vehicle	<b>0</b>	<b>0</b>	<b>0</b>	852	211	–	1,725	0
Wine	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	14	0
Zoo	<b>0</b>	<b>0</b>	<b>0</b>	9	<b>0</b>	<b>0</b>	1	0
Mean	<b>0</b>	<b>0</b>	<b>0*</b>	259	10	33*	238	0

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

Table W15: 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 15% 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.