

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