

Dataset	20% CS						KMEANS
	PCCC	PCCC-R	COPKM	CSC	DILS	LCC	
Appendicitis	<b>1.00</b>	<b>1.00</b>	–	<b>1.00</b>	<b>1.00</b>	0.57	0.33
Breast Cancer	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.00	<b>1.00</b>	–	0.67
Bupa	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	-0.01	<b>1.00</b>	<b>1.00</b>	-0.00
Circles	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	–	-0.00
Ecoli	<b>0.98</b>	<b>0.98</b>	–	0.32	0.81	0.80	0.39
Glass	<b>0.93</b>	<b>0.93</b>	–	0.12	0.79	0.74	0.24
Haberman	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.01	<b>1.00</b>	<b>1.00</b>	0.09
Hayesroth	<b>0.92</b>	<b>0.92</b>	–	0.23	<b>0.92</b>	0.78	0.06
Heart	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.93	<b>1.00</b>	<b>1.00</b>	0.41
Ionosphere	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.05	<b>1.00</b>	–	0.17
Iris	<b>0.64</b>	<b>0.64</b>	0.60	0.56	0.57	0.63	0.63
Led7Digit	<b>0.99</b>	<b>0.99</b>	–	0.05	0.13	0.50	0.43
Monk2	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.89	<b>1.00</b>	–	0.05
Moons	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.96	<b>1.00</b>	<b>1.00</b>	0.47
Movement Libras	0.53	<b>0.58</b>	–	0.03	0.10	0.31	0.31
Newthyroid	0.98	0.98	0.98	0.86	<b>1.00</b>	0.85	0.59
Saheart	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.01	<b>1.00</b>	–	0.07
Sonar	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	-0.00	<b>1.00</b>	<b>1.00</b>	0.00
Soybean	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.13	0.70	0.79	0.90
Spectfheart	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	-0.01	<b>1.00</b>	<b>1.00</b>	-0.10
Spiral	<b>1.00</b>	<b>1.00</b>	–	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.03
Tae	<b>0.96</b>	<b>0.96</b>	–	0.36	0.94	0.54	0.03
Vehicle	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.30	0.82	–	0.08
Wine	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>	0.73	0.61	0.90	0.85
Zoo	<b>0.93</b>	<b>0.93</b>	0.73	0.35	0.63	0.84	0.79
Mean	<b>0.95</b>	<b>0.95</b>	0.65*	0.39	0.84	0.61*	0.30

\*Nan values (–) are replaced with 0 before computing the mean.

Table W8: Average Adjusted Rand Index (ARI) values of the PCCC and the PCCC-R algorithms and the four state-of-the-art algorithms (COPKM, CSC, DILS, LCC) obtained with constraint sets of size 20% CS. Higher values indicate more overlap with the ground truth assignment. The highest values are stated in bold. The column KMEANS reports the average ARI values that were obtained with the unconstrained k-means algorithm. The hyphen indicates that the respective algorithm returned no solution within the time limit of 1,800 seconds.