

Dataset	Objects	Features	Clusters	0.5% CS						KMEANS
				PCCC-N2-S	PCCC-N2-S-RD	COPKM	LCC	CSC	DILS	
Banana5,300		2	2	0.014	0.014	–	0.006	0.000	0.003	0.01
Letter20,000		16	26	0.149	0.153	0.156	0.148	–	0.000	0.14
Shuttle57,999		9	7	0.647	0.654	–	0.312	–	–	0.41
CIFAR100,000		3,072	10	0.043	0.043	–	–	–	–	0.04
CIFAR100										
CIFAR100,000		3,072	100	0.022	0.022	–	–	–	–	0.02
MNIST10,000		784	10	0.372	0.374	0.217	–	–	–	0.31
Mean				0.208	0.210	0.062*	0.078*	0.000*	0.000*	0.15

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

Table W101: 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 0.5% 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.