

| Dataset | Objects | Features | Clusters | 1% CS | | | | | | | |
|-----------|---------|----------|----------|--------------|--------------|--------------|--------|--------------|---------|--------|--------|
| | | | | PCCC-N2-S | PCCC-N5-S | PCCC-N2-S-RD | COPKM | LCC | CSC | DILS | KMEANS |
| Banana | 5,300 | 2 | 2 | 0.015 | 0.015 | 0.015 | – | 0.011 | -0.000 | 0.004 | 0.017 |
| Letter | 20,000 | 16 | 26 | 0.162 | 0.162 | 0.159 | 0.154 | 0.162 | – | 0.000 | 0.149 |
| Shuttle | 57,999 | 9 | 7 | 0.956 | 0.965 | 0.932 | – | – | – | – | 0.411 |
| CIFAR 10 | 60,000 | 3,072 | 10 | 0.055 | 0.033 | 0.058 | – | – | – | – | 0.040 |
| CIFAR 100 | 60,000 | 3,072 | 100 | 0.021 | 0.021 | 0.021 | – | – | – | – | 0.021 |
| MNIST | 70,000 | 784 | 10 | 0.546 | 0.241 | 0.683 | – | – | – | – | 0.312 |
| Mean | | | | 0.292 | 0.239 | 0.311 | 0.026* | 0.029* | -0.000* | 0.001* | 0.159 |

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

Table W102: Average Adjusted Rand Index (ARI) values of the versions of the PCCC algorithm and the four state-of-the-art algorithms (COPKM, CSC, DILS, LCC) obtained with constraint sets of size 1% 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 3.600 seconds.