

Additional files 3: Table S1

Table 1 Table S1: The performance of clustering methods on eleven real UCI datasets measured by Purity. Purity is defined as the proportion of correctly clustered samples (with respect to the groups in the 'ground truth').

Dataset	Purity						MS CkNN
	c fixed to ground truth						
	k-means	Gaussian mixture	Hierarchical	Spectral N-cut	Spectral NJW	MS CkNN	
Iris	0.8933	0.9800	0.8400	0.9000	0.9067	0.9000	0.9000
Glass	0.5467	0.5280	0.4112	0.5327	0.5748	0.6636	0.6589
Wine	0.9663	0.9775	0.8371	0.9831	0.9831	0.9494	0.9494
WBDC	0.9104	0.9016	0.6309	0.9086	0.9121	0.9543	0.9543
Control chart	0.7000	0.5333	0.6117	0.8133	0.8117	0.7983	0.6667
Parkinson	0.7538	0.7538	0.7538	0.7538	0.7538	0.7897	0.8564
Vertebral	0.6774	0.7774	0.4839	0.7065	0.6968	0.7613	0.7613
Breast tissue	0.5472	0.6038	0.4717	0.6132	0.6321	0.6132	0.5283
Seeds	0.9190	0.9333	0.8762	0.9000	0.9000	0.9048	0.9048
Image Seg.	0.5541	0.5545	0.1671	0.5926	0.6541	0.6082	0.7649
Yeast	0.5492	0.4623	0.3908	0.5303	0.5283	0.5364	0.5364
Average	0.7289	0.7278	0.5886	0.7486	0.7594	0.7708	0.7710