Online Appendix for "Benchmarking Minimax Linkage in Hierarchical Clustering"

Xiao Hui Tai and Kayla Frisoli

The following are full evaluation results for all data sets and evaluation metrics, using true k and all k. Tables A.1 and A.2 report the results for all linkage types considered, for the true k of each data set. Figures A.1 through A.12 show the distribution of the metrics across all possible values of k.

Table A.1 Results for true k. Entries in boldface are the linkage types with best performance for that metric.

Data set (k = truth)	Linkage type	Max minimax radius	Misclassification	Precision	Recall
Olivetti Faces ($k = 40$)	single	3394.93	0.40	0.04	0.78
	complete	2606.25	**0.04**	**0.31**	0.49
	average	2449.69	0.07	0.18	0.60
	centroid	3259.74	0.79	0.02	**0.83**
	minimax	**2293.45**	0.05	0.24	0.57
Colon Cancer ($k = 2$)	single	0.34	**0.46**	**0.54**	**0.98**
	complete	**0.28**	0.48	0.53	0.87
	average	**0.28**	0.48	0.53	0.87
	centroid	**0.28**	0.47	0.53	0.90
	minimax	0.29	0.48	0.53	0.92
Prostate Cancer ($k = 2$)	single	0.48	0.50	0.50	**0.98**
	complete	**0.33**	0.49	0.50	0.77
	average	0.35	**0.49**	0.50	0.73
	centroid	0.40	**0.49**	0.50	0.69
	minimax	0.35	**0.49**	0.50	0.76
Spherical-ℓ_2 ($k = 3$)	single	6.07	0.66	0.33	**0.99**
	complete	**5.13**	**0.24**	**0.63**	0.64
	average	5.95	0.66	0.33	0.98
	centroid	6.07	0.66	0.33	**0.99**
	minimax	5.35	0.25	0.62	0.65
Spherical-ℓ_1 ($k = 3$)	single	15.97	0.66	0.33	**0.99**
	complete	**14.26**	**0.33**	**0.51**	0.55
	average	15.72	0.66	0.33	0.98
	centroid	15.75	0.66	0.33	**0.99**
	minimax	14.87	**0.33**	**0.51**	0.51
Elliptical-ℓ_2 ($k = 3$)	single	6.79	0.66	0.33	**0.99**
	complete	**5.95**	**0.35**	**0.48**	0.51
	average	6.66	0.66	0.33	0.96
	centroid	6.76	0.66	0.33	**0.99**
	minimax	6.21	0.38	0.44	0.51

Xiao Hui Tai
University of California, Berkeley, e-mail: xtai@berkeley.edu

Kayla Frisoli
Carnegie Mellon University e-mail: kfrisoli@andrew.cmu.edu

Table A.2 Results for true k (cont.)

Data set	Linkage type	Max minimax radius	Misclassification	Precision	Recall
Elliptical-ℓ_1 ($k = 3$)	single	17.40	0.66	0.33	**0.99**
	complete	16.40	0.38	0.44	0.59
	average	17.40	0.66	0.33	0.96
	centroid	17.37	0.66	0.33	**0.99**
	minimax	**15.60**	**0.33**	**0.50**	0.57
Outliers-ℓ_2 ($k = 3$)	single	6.46	0.66	0.33	**0.99**
	complete	**5.81**	0.46	0.38	0.65
	average	6.12	0.65	0.33	0.95
	centroid	6.37	0.66	0.33	0.98
	minimax	5.95	**0.39**	**0.44**	0.65
Outliers-ℓ_1 ($k = 3$)	single	17.39	0.66	0.33	**0.99**
	complete	15.99	0.42	0.39	0.50
	average	16.37	0.66	0.33	0.97
	centroid	16.37	0.66	0.33	0.98
	minimax	**14.79**	**0.26**	**0.60**	0.61
Iris ($k = 3$)	single	2.97	0.23	0.59	**0.99**
	complete	2.19	0.20	0.67	0.79
	average	2.56	0.22	0.60	0.96
	centroid	2.97	0.23	0.59	**0.99**
	minimax	**2.09**	**0.17**	**0.71**	0.79
NBIDE ($k = 12$)	single	0.82	0.23	0.23	0.89
	complete	0.77	0.05	0.66	0.79
	average	0.75	0.03	0.77	0.91
	centroid	0.83	0.80	0.08	0.87
	minimax	**0.73**	**0.02**	**0.84**	**0.92**
FBISW ($k = 69$)	single	0.75	0.01	0.33	0.86
	complete	0.65	**0.00**	**0.83**	**0.93**
	average	0.63	**0.00**	0.77	0.91
	centroid	0.82	0.17	0.02	0.58
	minimax	**0.59**	**0.00**	0.70	0.90

Fig. A.1 Results for Olivetti faces

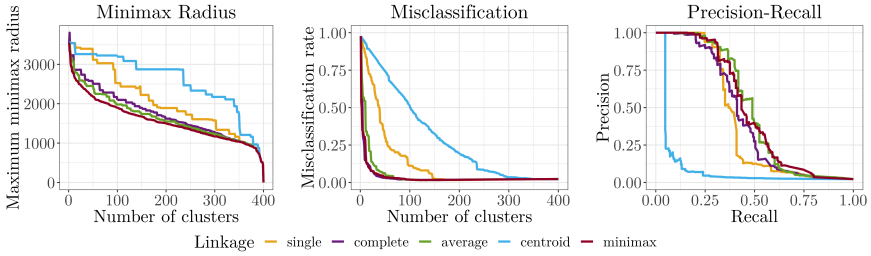

Fig. A.2 Results for Colon Cancer

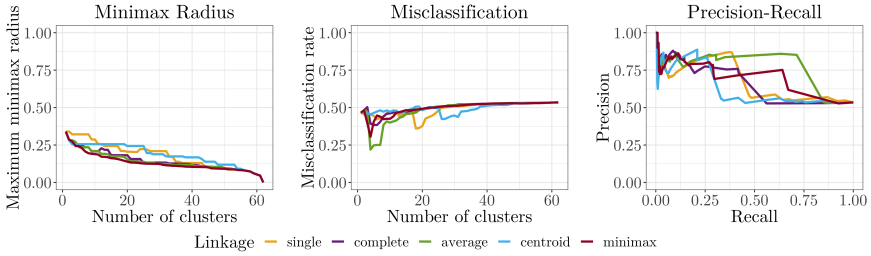

Fig. A.3 Results for Prostate Cancer

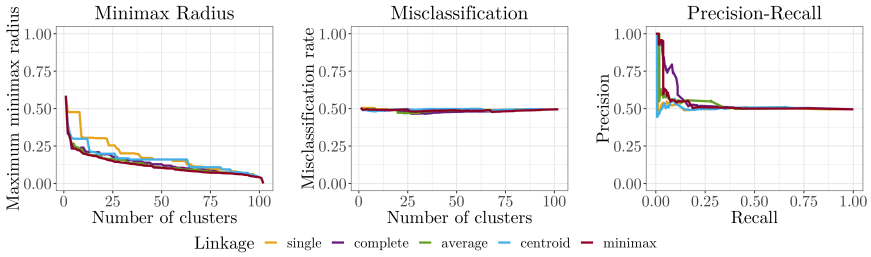

Fig. A.4 Results for simulation: spherical-ℓ_2

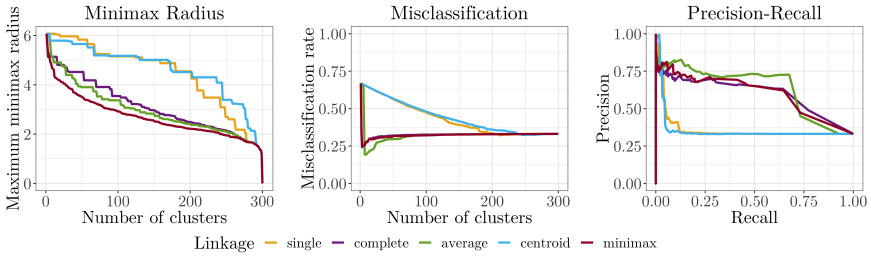

Fig. A.5 Results for simulation: spherical-ℓ_1

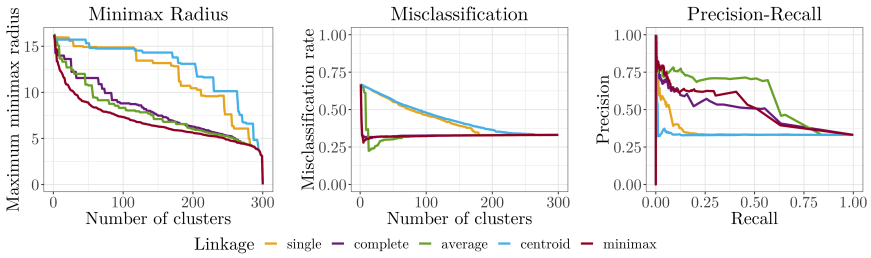

Fig. A.6 Results for simulation: elliptical-ℓ_2

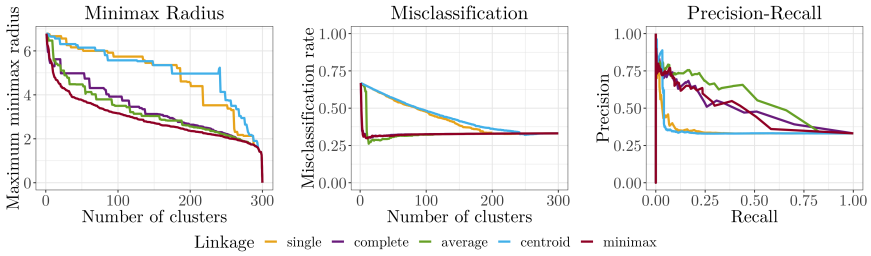

Fig. A.7 Results for simulation: elliptical-ℓ_1

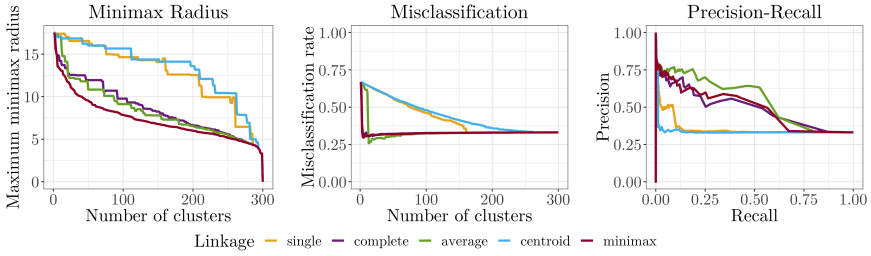

Fig. A.8 Results for simulation: outliers-ℓ_2

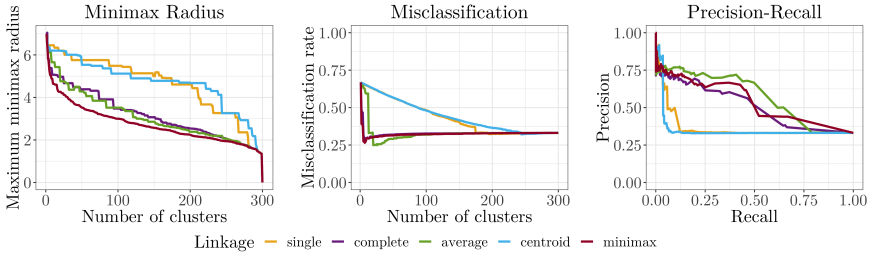

Fig. A.9 Results for simulation: outliers-ℓ_1

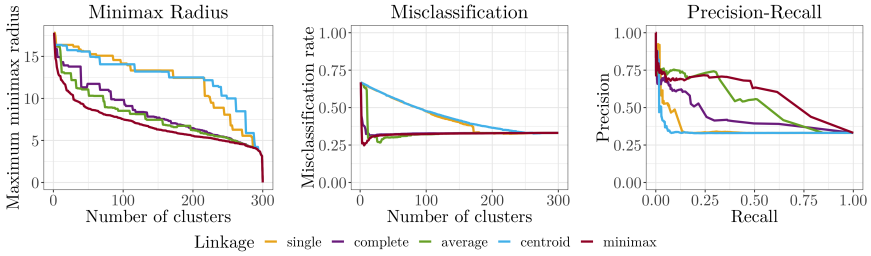

Fig. A.10 Results for iris

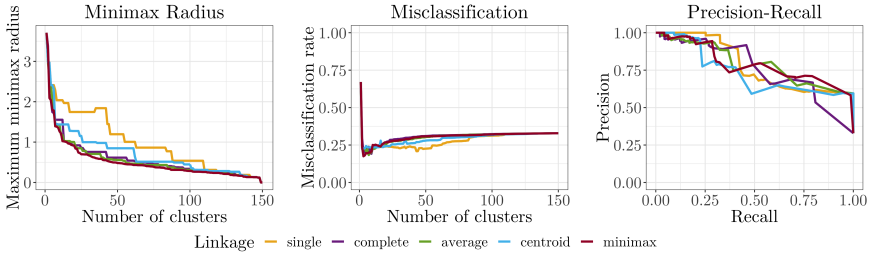

Fig. A.11 Results for NBIDE study

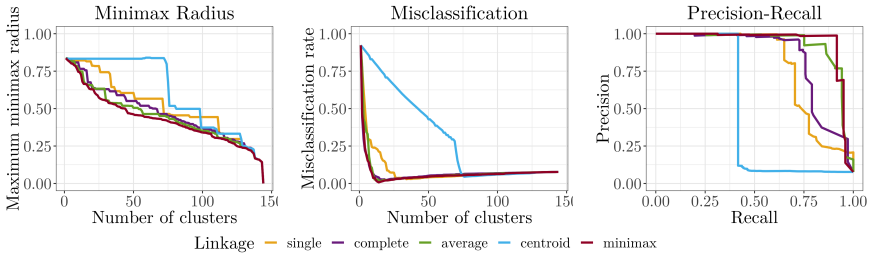

Fig. A.12 Results for FBI S&W study