

Table 1: Runtime, Pearson’s correlation coefficient between distance matrices and Fowlkes-Mallows index for $k = 4$ and $k = 8$. The ‘reference’ method calculates distances from the original sequences. We show only two assembly algorithms that gave the highest correlation and the better algorithm of pairs MES/MESS and MESSG/MESSGM.

Dataset	method	finished	<u>assem.</u> ms	<u>distances</u> ms	<u>UPGMA</u> ms	<u>NJ</u> ms	corr.	<u>UPGMA</u> B_4	<u>UPGMA</u> B_8	<u>NJ</u> B_4	<u>NJ</u> B_8
Influenza	reference	112/112	0	4,893.33	4.13	11.29	1	1	1	1	1
	$\max(R_A , R_B)$	112/112	0	186.87	0.58	0.96	.801	.67	.32	.66	.32
	Dist _{MES}	106/112	0	857,243.98	0.22	0.84	.824	.33	.69	.52	.82
	Dist _{MESS}	106/112	0	854,821.17	0.63	0.41	.942	1	.85	1	.83
	Dist _{MESSG}	96/112	0	859,725.65	0.23	0.33	.98	.99	1	1	.99
	Dist _{MESSGM}	96/112	0	862,413.18	0.25	0.32	.978	.99	1	1	.98
	Dist _{MESSGq}	112/112	0	87,278.04	0.22	0.42	.976	1	.99	1	.99
	ABySS	79/112	20,839.76	9,623.81	0.22	0.25	.637	.41	.49	.54	.48
	Edena	55/112	367	10,757.31	0.24	0.44	.641	.43	.52	.54	.53
	SSAKE	51/112	2,983.75	12,352.33	0.25	0.25	.72	.46	.59	.55	.66
	Velvet	107/112	463.48	14,329.37	0.25	0.34	.705	.38	.64	.55	.78
	SPAdes	31/112	12,438.81	13,684	0.16	0.23	.722	.56	.68	.66	.72
Various	reference	112/112	0	71,019.2	4.81	11.1	1	1	1	1	1
	$\max(R_A , R_B)$	112/112	0	441.49	0.62	1.16	.901	.67	.66	.85	.92
	Dist _{MES}	72/112	0	1,214,293.61	0.32	0.53	.589	.39	.41	.47	.32
	Dist _{MESS}	72/112	0	1,212,284.96	0.33	0.53	.928	.63	.79	.87	.98
	Dist _{MESSG}	64/112	0	1,480,700.73	0.34	0.5	.934	.63	.88	.91	.96
	Dist _{MESSGM}	64/112	0	1,482,902.92	0.3	0.56	.934	.65	.89	.87	.94
	Dist _{MESSGq}	110/112	0	660,983.1	0.4	0.54	.929	.66	.76	.84	.98
	ABySS	69/112	27,778.49	115,247.36	0.28	0.54	.394	.43	.17	.43	.24
	Edena	57/112	1,112.84	156,249.28	0.39	0.61	.447	.47	.21	.48	.27
	SSAKE	57/112	8,471.35	160,368.25	0.35	0.63	.503	.47	.28	.49	.32
	Velvet	109/112	1,726.28	165,184.43	0.36	0.61	.39	.46	.19	.48	.31
	SPAdes	37/112	15,978.65	167,090.27	0.27	0.54	.549	.52	.36	.55	.35
Hepatitis	reference	9/9	0	2,300,822.78	33.22	51.89	1	1	1	1	1
	$\max(R_A , R_B)$	9/9	0	8,301.22	7.78	18.44	.181	.55	.37	.72	.83
	Dist _{MES}	9/9	0	12,699,975.67	37.89	82.44	.833	1	.95	1	.92
	Dist _{MESS}	9/9	0	12,566,601.33	399.44	274.44	.788	.9	.95	1	.94
	Dist _{MESSG}	9/9	0	26,958,877.56	376.22	606.67	.964	1	.95	1	.92
	Dist _{MESSGM}	9/9	0	26,943,068	134.56	207.22	.957	1	.95	.98	.97
	Dist _{MESSGq}	9/9	0	616,507.11	3.78	7.89	.901	.94	.96	1	.94
	ABySS	9/9	151,744	2,022,804.67	53.22	72.44	.428	.73	.74	.93	.59
	Edena	9/9	10,546.78	2,052,872.89	1.44	4.44	.507	.82	.79	.95	.79
	SSAKE	9/9	91,991.33	2,327,985.22	1.44	5	.554	.8	.71	.94	.78
	Velvet	9/9	22,599.89	3,016,592	1.67	4.11	.783	.8	.84	.96	.86
	SPAdes	9/9	97,626.44	1,882,311.56	304.89	119.11	.578	.82	.83	.95	.83
Chromosomes	reference	1/1	0	886,125	7	20	1	1	1	1	1
	$\max(R_A , R_B)$	1/1	0	868	1	2	.331	.64	.4	.61	.3
	Dist _{MES}	1/1	0	14,006,230	0	1	.886	.42	.26	.6	.28
	Dist _{MESS}	1/1	0	14,107,590	1	0	.359	.6	.44	.61	.3
	Dist _{MESSG}	1/1	0	29,676,206	0	1	.364	.54	.47	.64	.35
	Dist _{MESSGM}	1/1	0	29,739,506	1	2	.313	.54	.56	.64	.31
	Dist _{MESSGq}	1/1	0	219,485	0	2	.366	.53	.47	.64	.35
	ABySS	1/1	45,123	387,886	0	1	.267	.52	.17	.56	.27
	Edena	0/1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	SSAKE	1/1	57,584	70,383	1	1	.718	.47	.19	.81	.15
	Velvet	1/1	6,387	221,633	0	1	.568	.48	.25	.8	.36
	SPAdes	1/1	27,009	125,688	1	2	.444	.41	.18	.58	.23

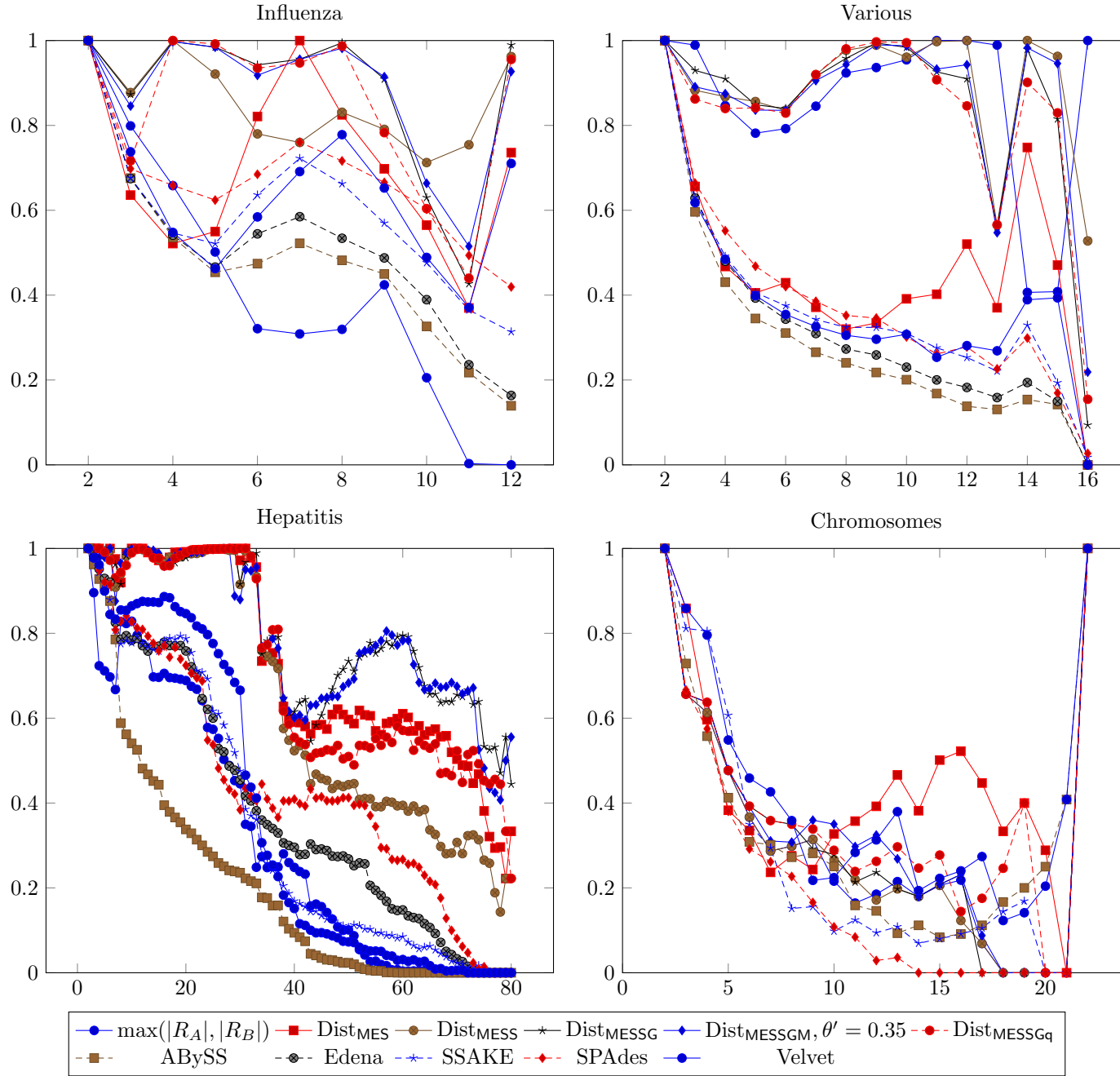


Figure 1: Plots of Fowlkes-Mallows index B_k versus k . The index compares trees generated by the neighbor-joining algorithm. The tree is compared with the tree generated from the original sequences. If all values are equal to 1, the structures of the trees are the same.

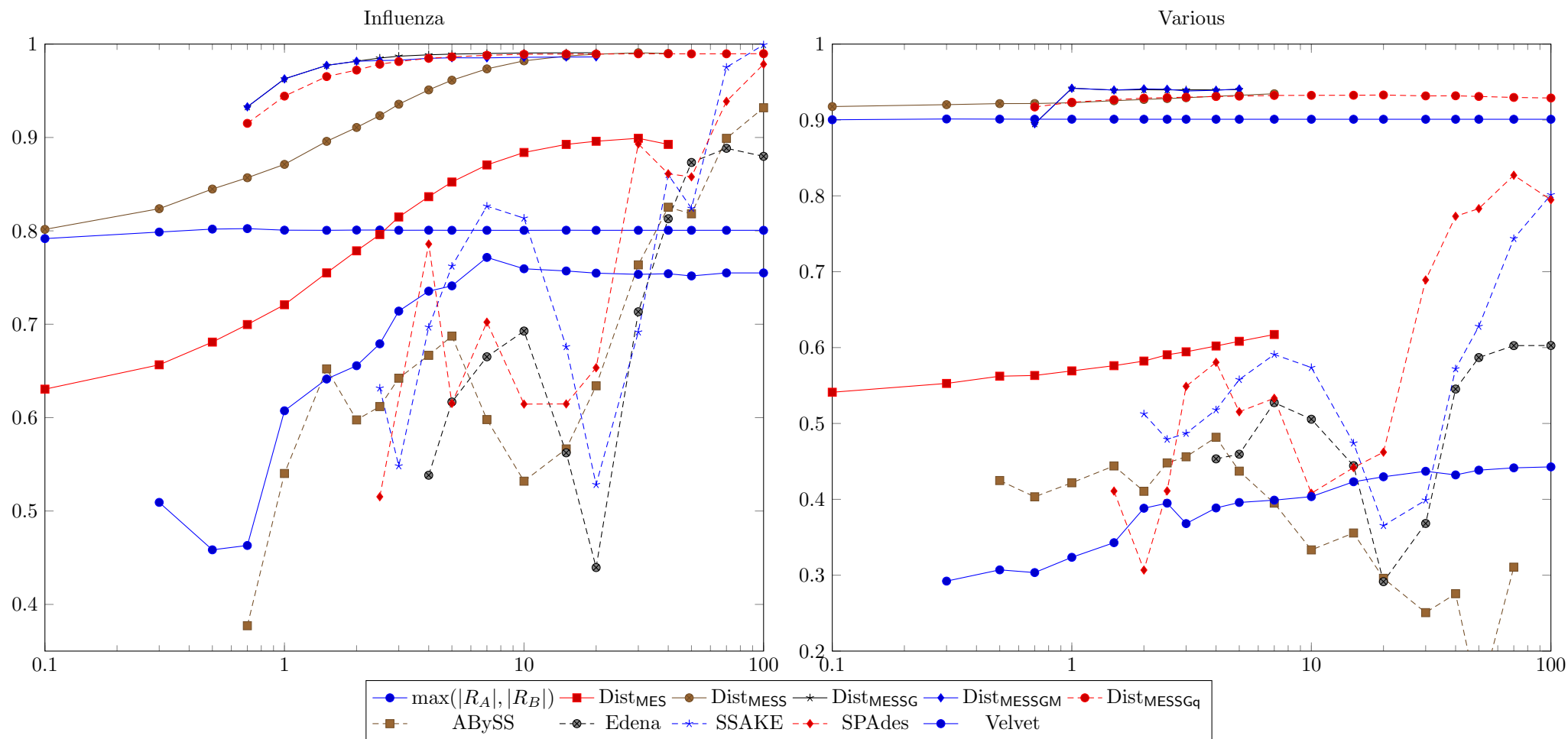


Figure 2: Plot of average Pearson's correlation coefficient for several choices of coverage values.

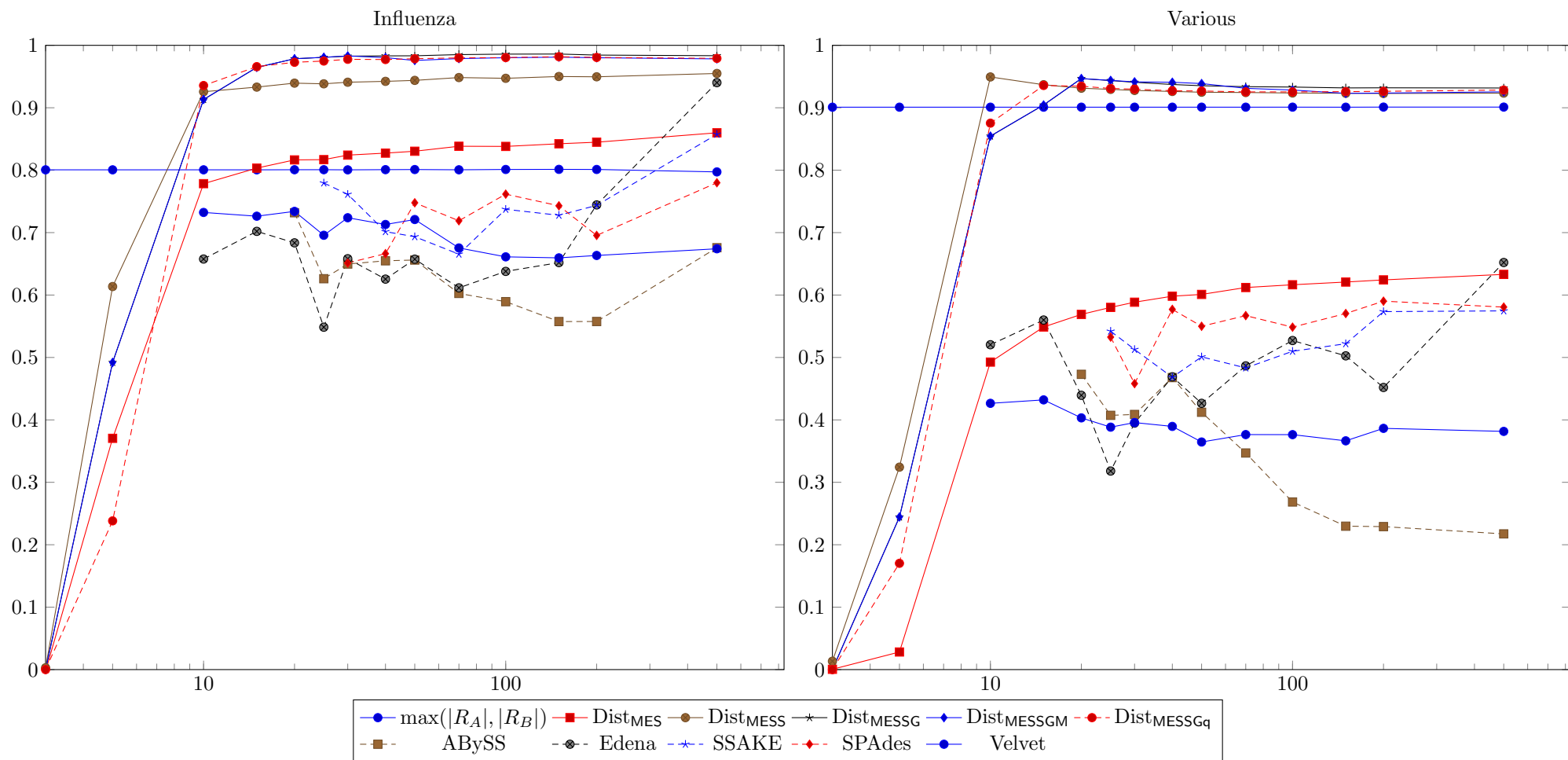


Figure 3: Plot of average Pearson's correlation coefficient for several choices of read length.