Table 1: Average 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 assembly algorithms that gave the highest and the lowest correlation.

Dataset	method	finished	assem. ms	distances ms	corr.	NJ B_4	NJ B_8
Influenza	reference	112/112	0	2,517.55	1	1	1
	$\max(\|R_A\|, \|R_B\|)$	112/112	0	183.57	.801	.66	.32
	$\text{Dist}_{\mathsf{MESSGq}}$	112/112	0	43,552.69	.966	1	.97
	longest contig ABySS	87/112	23,460.91	1,338.46	.67	.62	.43
	longest contig Edena	72/112	285.13	1,344.32	.675	.62	.46
	longest contig SSAKE	67/112	2,114.82	1,189.42	.655	.59	.37
	longest contig Velvet	110/112	391.74	101.4	.569	.46	.23
	longest contig SPAdes	43/112	12,460.6	2,126.93	.751	.71	.56
	longest contig opt_{10}	112/112	0	50.87	.503	.46	.12
	longest contig opt_{20}	112/112	0	40.46	.502	.44	.13
	longest contig opt_{30}	112/112	0	19.87	.512	.45	.12
	longest contig opt_{40}	112/112	0	71.88	.513	.45	.13
	longest contig opt_{50}	112/112	0	252.14	.542	.47	.18
	longest contig opt_{60}	112/112	0	341.85	.544	.49	.19
	longest contig opt_{70}	112/112	0	558.96	.586	.53	.27
	longest contig opt_{80}	112/112	0	826.88	.609	.57	.34
	longest contig opt_{90}	112/112	0	1,208.46	.666	.63	.43
	$\text{Dist}(C_A, C_B)$ ABySS	87/112	23,460.91	17,652.57	.954	.98	.84
	$\text{Dist}(C_A, C_B)$ Edena	72/112	285.13	18,232.13	.962	.99	.86
	$\text{Dist}(C_A, C_B)$ SSAKE	67/112	2,114.82	17,483.33	.949	.98	.87
	$\text{Dist}(C_A, C_B)$ Velvet	110/112	391.74	22,256.65	.96	.99	.92
	$\text{Dist}(C_A, C_B)$ SPAdes	43/112	12,460.6	20,968.09	.975	.99	.95
	$\text{Dist}(C_A, C_B)$ opt_{10}	72/112	0	832.1	.834	.85	.45
	$\text{Dist}(C_A, C_B)$ opt_{20}	106/112	0	2,753.36	.863	.92	.54
	$\text{Dist}(C_A, C_B)$ opt_{30}	111/112	0	5,001.62	.896	.98	.63
	$\text{Dist}(C_A, C_B)$ opt_{40}	112/112	0	7,476.58	.924	1	.76
	$\text{Dist}(C_A, C_B)$ opt_{50}	112/112	0	11,159.66	.95	1	.89
	$\text{Dist}(C_A, C_B)$ opt_{60}	112/112	0	14,861.54	.971	1	.96
	$\text{Dist}(C_A, C_B)$ opt_{70}	112/112	0	16,683.14	.979	1	.99
	$\text{Dist}(C_A, C_B)$ opt_{80}	112/112	0	20,851.62	.986	1	.98
	$\text{Dist}(C_A, C_B)$ opt_{90}	112/112	0	22,238.89	.987	1	.98
Hepatitis	reference	9/9	0	2,145,103.56	1	1	1
	$\max(\|R_A\|, \|R_B\|)$	9/9	0	7,738.22	.181	.72	.83
	$\text{Dist}_{\mathsf{MESSGq}}$	9/9	0	701,725.56	.897	1	.98
	longest contig ABySS	9/9	175,222.78	1,527,714.22	.53	.95	.69
	longest contig Edena	9/9	11,010.67	1,525,751.67	.515	.92	.76
	longest contig SSAKE	9/9	96,446.33	1,557,304.22	.5	.93	.8
	longest contig Velvet	9/9	22,860	3,447.33	.234	.93	.54
	longest contig SPAdes	9/9	103,682.89	1,872,232.78	.591	.95	.84
	$\text{Dist}(C_A, C_B)$ ABySS	9/9	175,222.78	29,339,917.89	.949	1	.88
	$\text{Dist}(C_A, C_B)$ Edena	9/9	11,010.67	26,593,653.33	.946	1	.89
	$\text{Dist}(C_A, C_B)$ SSAKE	9/9	96,446.33	29,465,435.78	.916	1	.9
	$\text{Dist}(C_A, C_B)$ Velvet	9/9	22,860	28,186,784.11	.966	1	.98
	$\text{Dist}(C_A, C_B)$ SPAdes	9/9	103,682.89	23,734,958.22	.959	1	.88

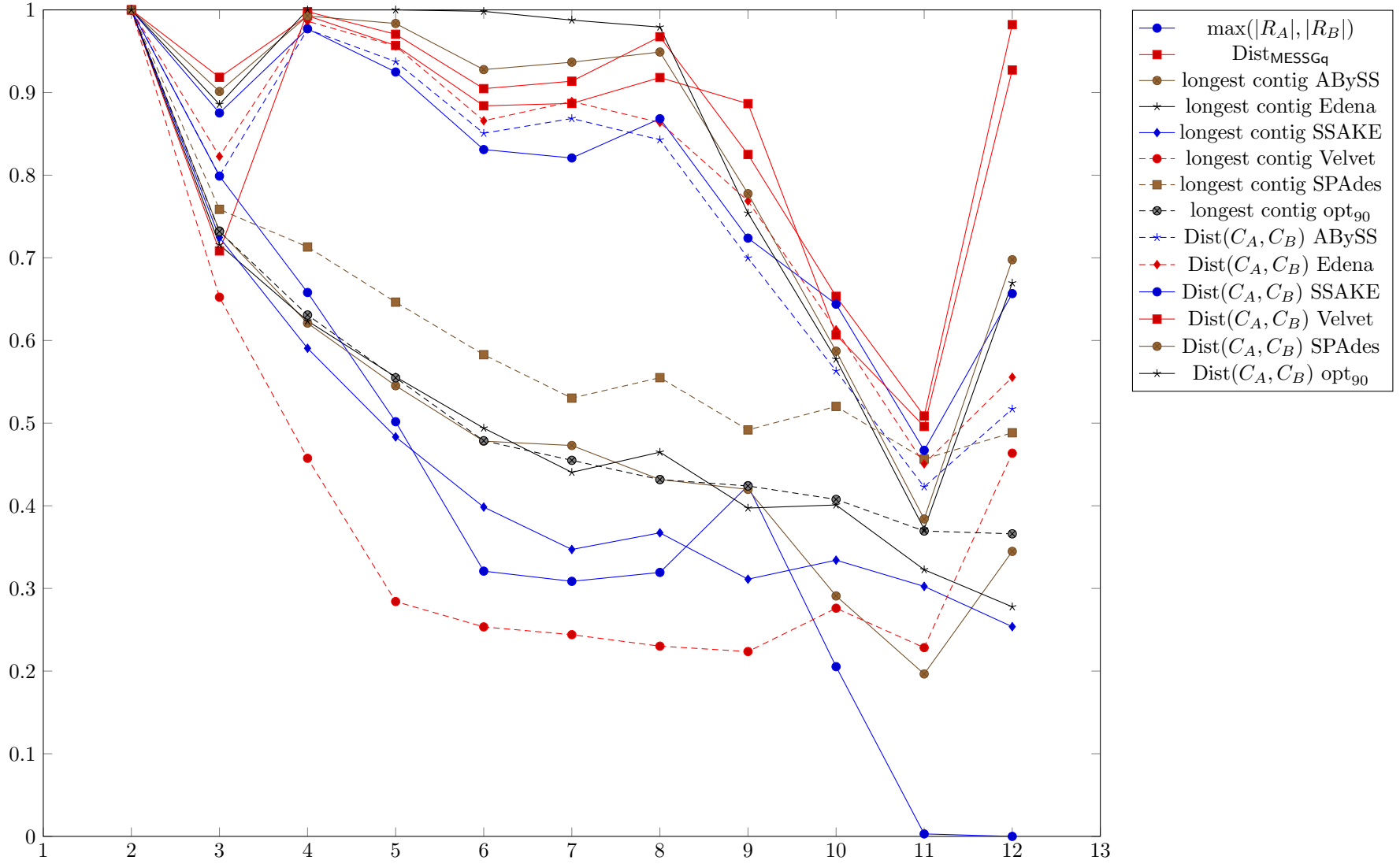

Figure 1: Plot of Fowlkes-Mallows index B_k versus $k$on influenza dataset. The index compares trees generated by the neighbor-joining algorithm. The tree is compared with the tree generated from the original sequences. If all values are equal to 1, the structures of the trees are the same.

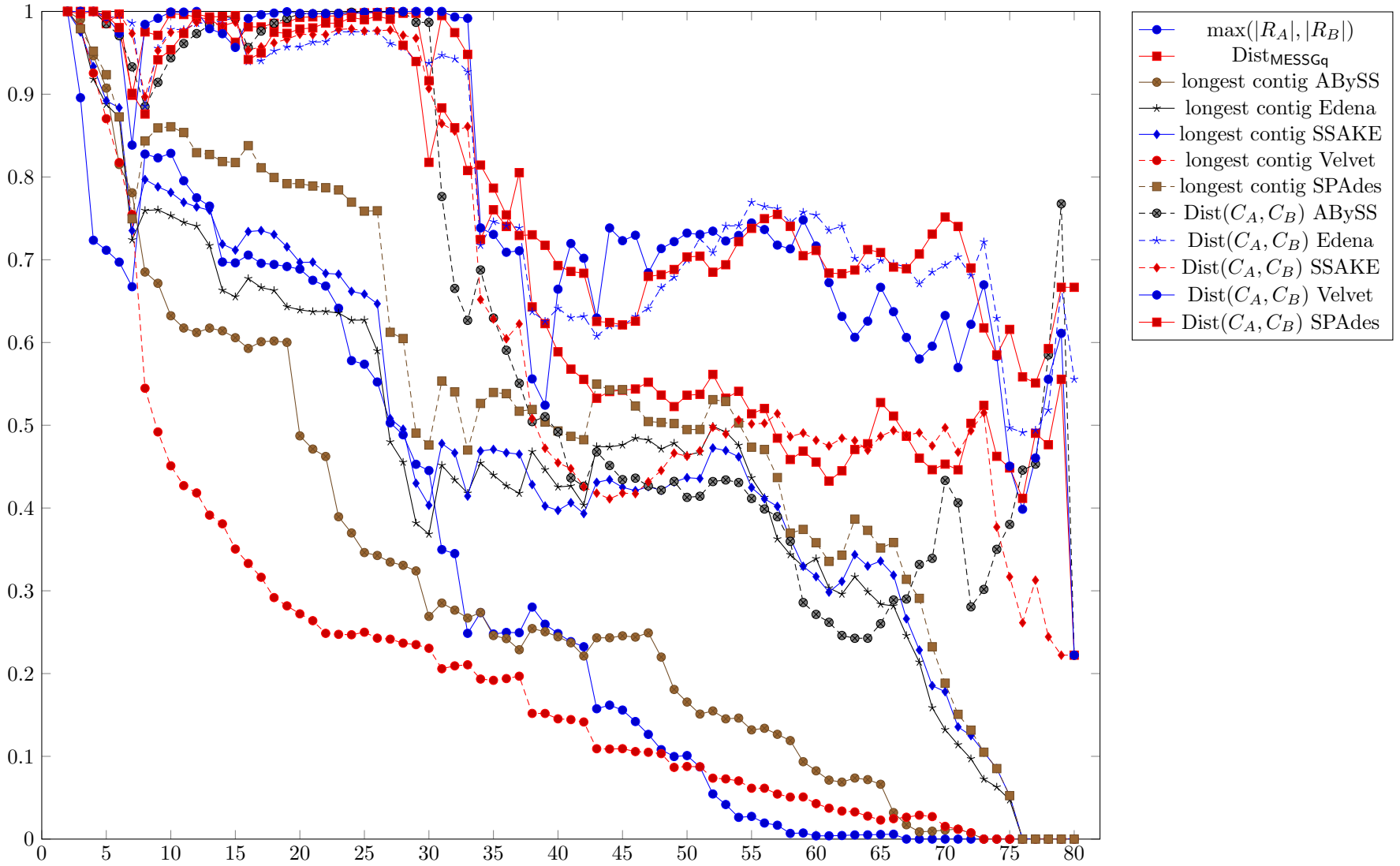

Figure 2: Plot of Fowlkes-Mallows index B_k versus k on hepatitis dataset. The index compares trees generated by the neighbor-joining algorithm.

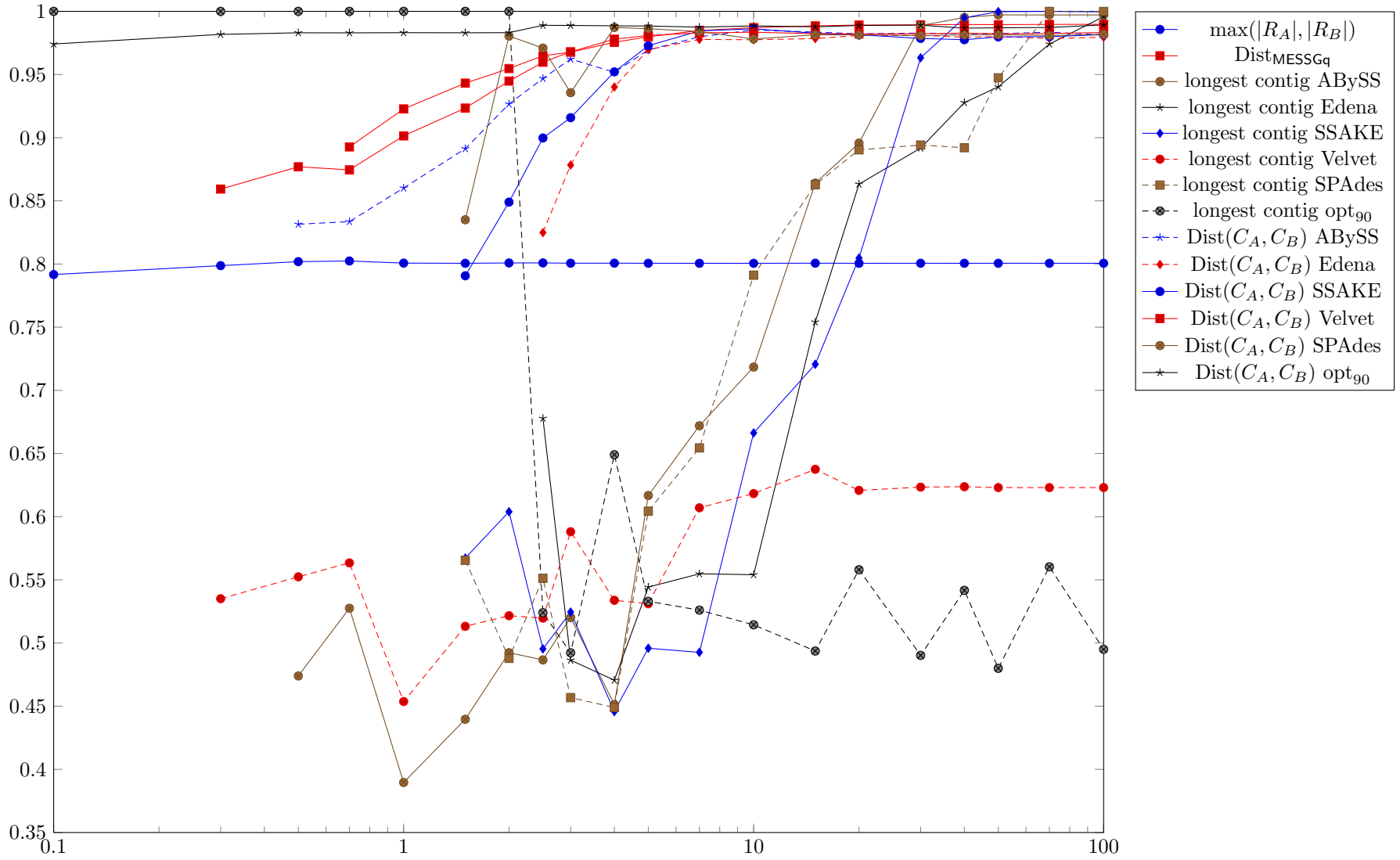

Figure 3: Plot of average Pearson's correlation coefficient for several choices of coverage values.

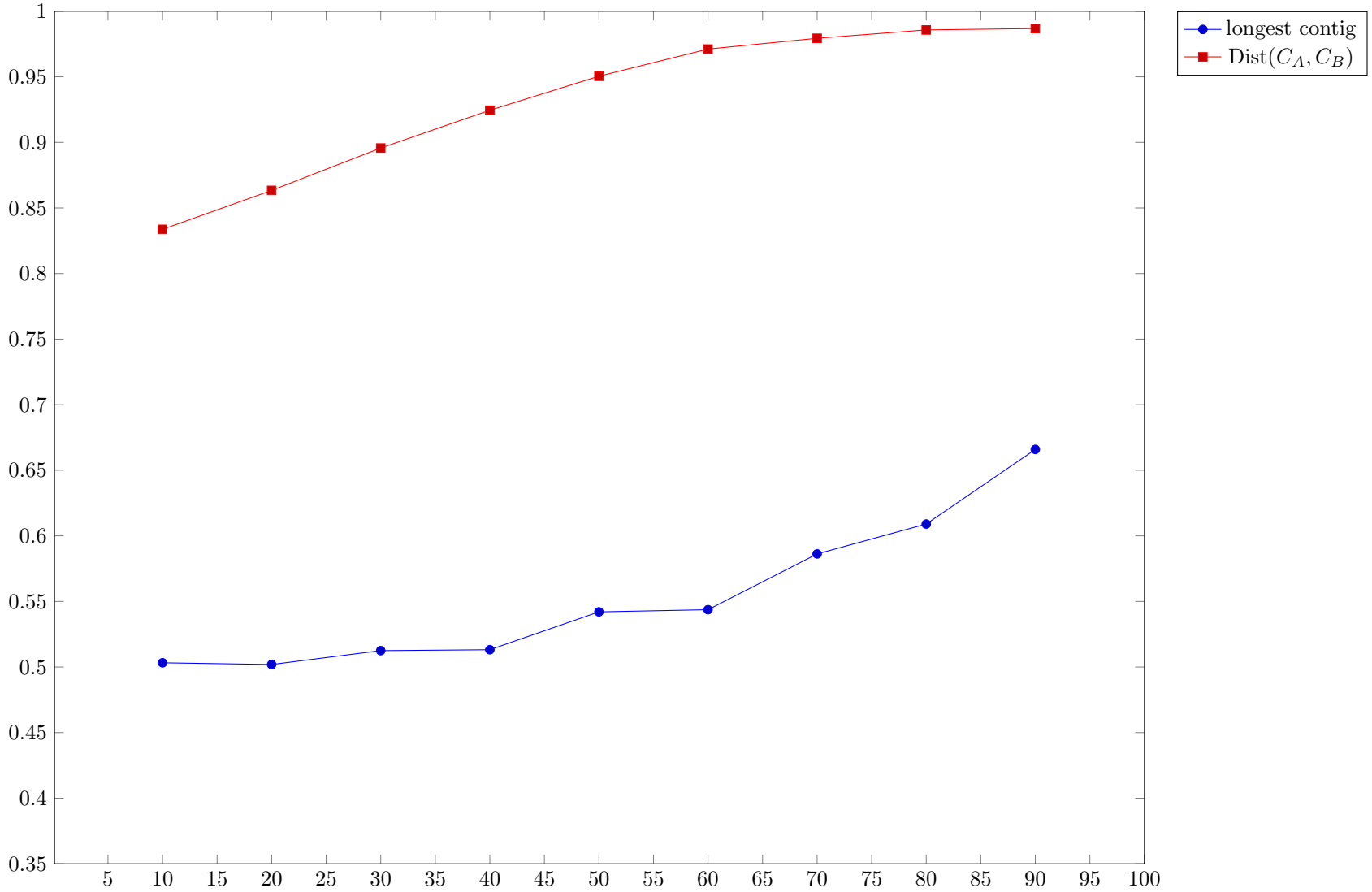

Figure 4: Plot of average Pearson's correlation coefficient on θ parameter for simulated assembly opt_θ on influenza dataset.