Evaluating String Distance Metrics for Reduction of Automatically Generated Test Suites

Author details suppressed for double-blind review

***Abstract*—This document contains all the analysis tables and figures for the experiments made for evaluating the diversity-based reduction.**

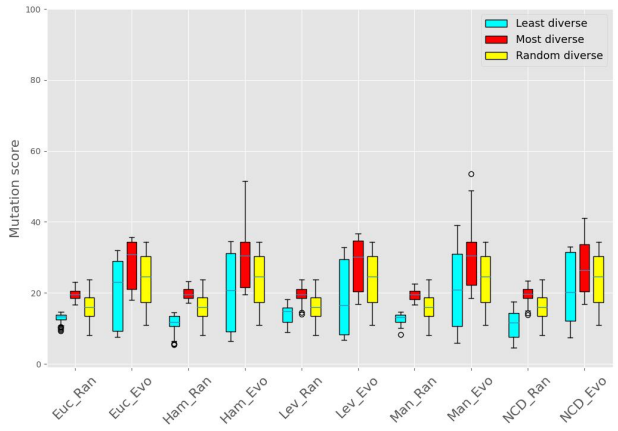
I. Results

In this document, we report the results of the experiments. We made our experiments on 13 test subjects taken from the "Defects4J" framework. For each test subject, we generated test suites using EvoSuite and Randoop, used five similarity metrics as a basis for reduction in addition to random reduction, used three different sample sizes for reduction, ran our experiments 30 times for random, least diverse, and most diverse test suites respectively, and made an evaluation based on mutation scores. In total, we ran 25740 experiments in this study. It took around 2436 hours to run these experiments. Also, we ran another 1100 experiments using different regression test suites in the smaller study to verify our findings. Table **??** presents all 13 test subjects with information about the original test suites' sizes and mutation score generated from both Randoop and EvoSuite given 3000 seconds and an upper size limit of 1500. The first column shows the project ID as defined in the Defects4J framework. The second and third columns display the test suite (TS) size and test suite mutation score (MS) for the Randoop test suites, respectively. The fourth and fifth columns display the test suite size and test suite mutation score for the EvoSuite test suites, respectively.

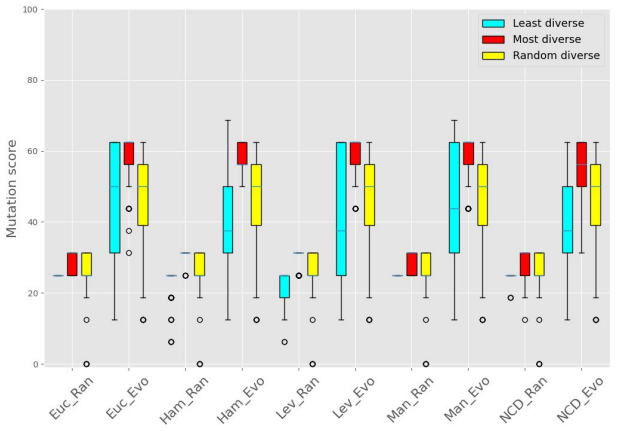
Figures 1 and 2 show the box plots of all the analysis results for all projects. Tables I-XIII show the analysis results of the projects. In each table, the first column is the used similarity metric (SM), and the second column is the size of the reduced test suite (TS). The third, fourth, and fifth columns display the average (Avg) and standard deviation (SD) for the mutation scores using both Randoop (Rand) and EvoSuite (Evo) for least diverse, most diverse, and random reductions, respectively. The sixth column displays the p-values of the statistical testing using both tools to show the statistical significance between LR (least diverse and random), and MR (most diverse and random). We used the *The Mann-Whitney U test* as our statistiscal test because the two samples (MR or LR) are independent and continues satisfying the preconditions of the Mann-Whitney U test. An α represents a number lower than 0.001, which is in the 99% significance level. The last column displays the *A12 effect size* proposed by Vargha and Delaney [1] using both tools between LR and MR. Effect size informs you how meaningful the relationship between LR and MR. The values in the effect size are in $[0, 1]$, where higher values (> 0.8) means large effect, values around 0.5 means medium effect, and small values (< 0.2) means small effect. Figures 1 and 2 show box plots of the achieved mutation scores for least diverse, most diverse, and random reduced test suites.

References

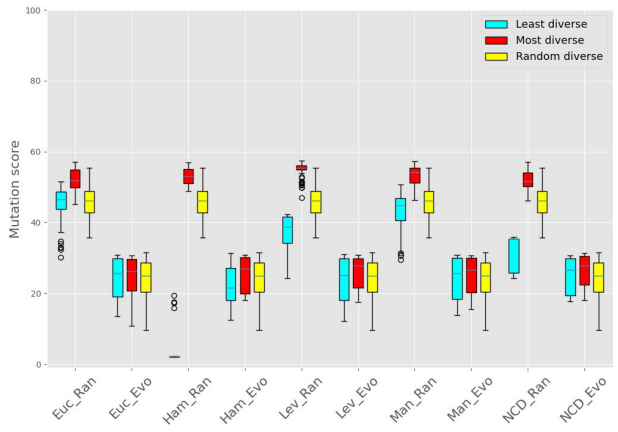
[1] A. Vargha and H. D. Delaney. A critique and improvement of the cl common language effect size statistics of mcgraw and wong. *Journal of Educational and Behavioral Statistics*, 25(2):101–132, 2000.



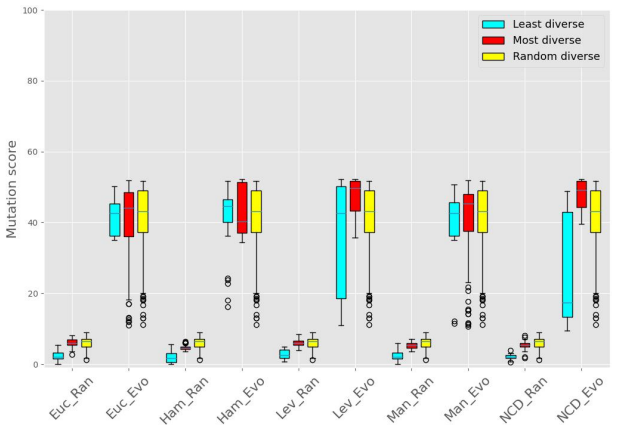
(a) Analysis on the "`Chart`" project.



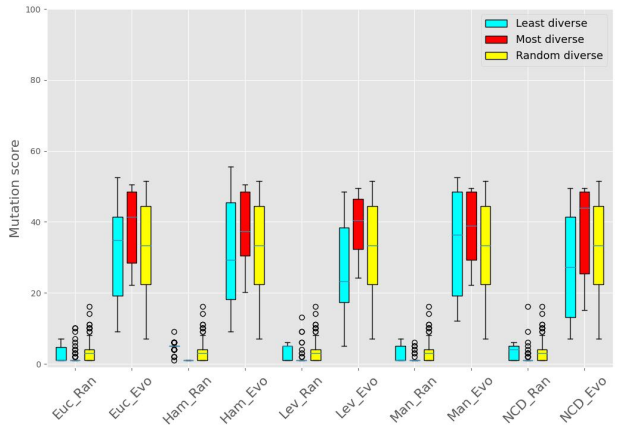
(b) Analysis on the "`Cli`" project.



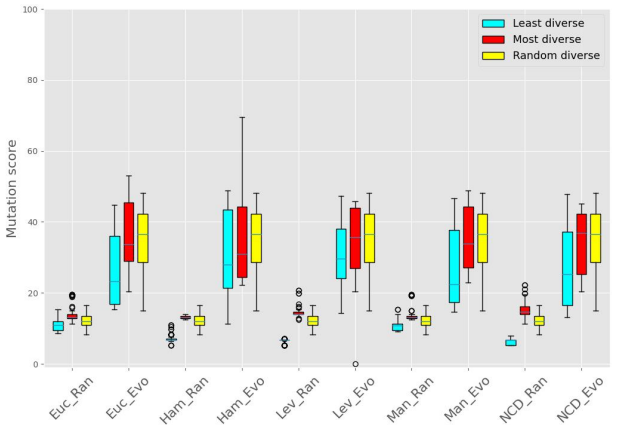
(c) Analysis on the "`Codec`" project.



(d) Analysis on the "`Compress`" project.

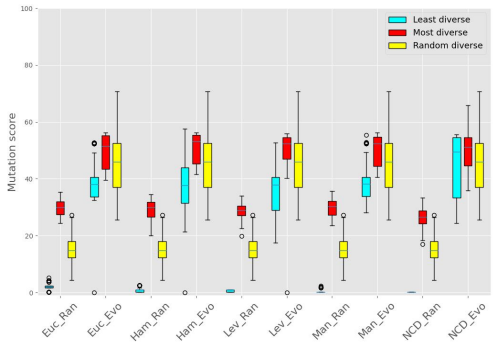


(e) Analysis on the "`Csv`" project.

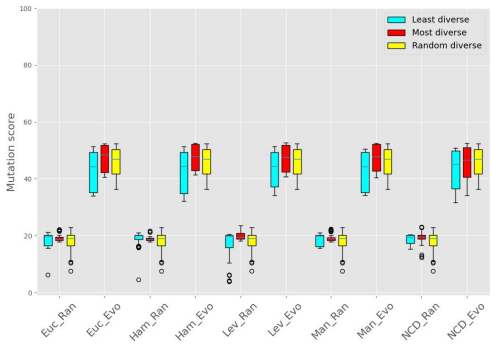


(f) Analysis on the "`Gson`" project.

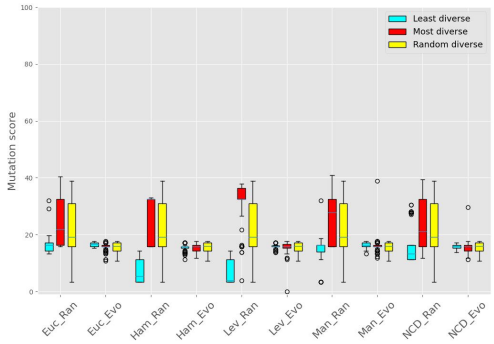
Fig. 1: The reduction analysis for "`Chart`", "`Chart`", "`Chart`", "`Chart`", "`Chart`", and "`Gson`".



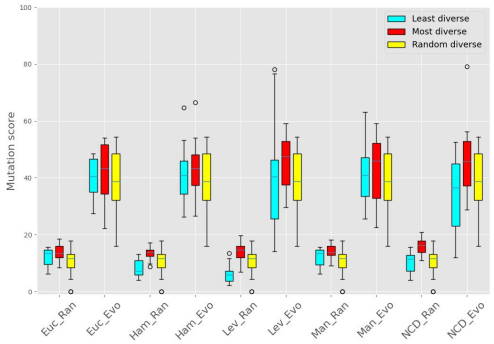
(a) Analysis on the "JacksonCore" project.



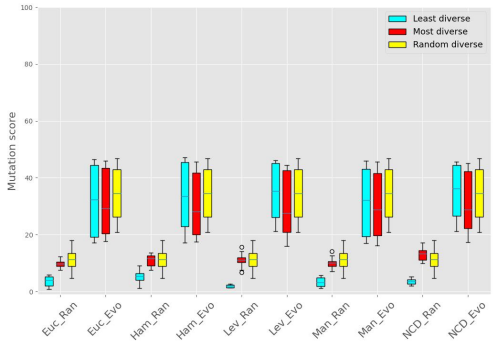
(b) Analysis on the "JacksonDatabind" project.



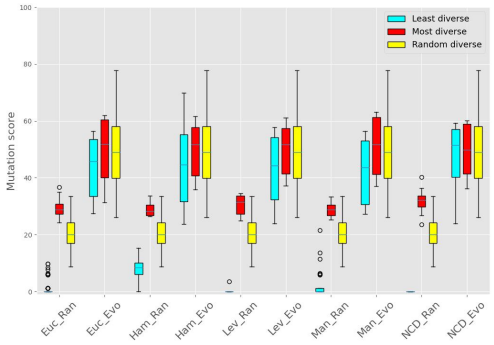
(c) Analysis on the "Jsoup" project.



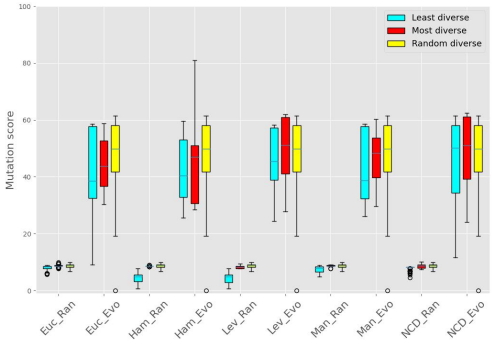
(d) Analysis on the "JxPath" project.



(e) Analysis on the "Lang" project.



(f) Analysis on the "Math" project.



(g) Analysis on the "Time" project.

Fig. 2: The reduction analysis for "JacksonCore", "JacksonDatabind", "Jsoup", "JxPath", "Lang", "Math", and "Time".

TABLE I: The reduction analysis for the Chart project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	37	11.6	0.96	8.9	1.6	17.9	0.73	20.5	1.2	12.5	2.28	16.1	2.5	0.055	α	α	α	0.36	0.99	0.01	0.92
	63	13.2	0.58	22.8	1.13	20.0	0.73	30.8	0.34	16.5	2.00	24.1	2.18	α	α	α	α	0.06	0.96	0.26	1.00
	90	14.2	0.49	30.6	1.14	20.8	0.83	34.7	0.46	18.6	2.36	31.3	1.72	α	α	α	α	0.05	0.86	0.41	0.99
Ham	37	10.0	2.63	9.4	2.52	18.3	0.54	21.0	0.67	12.5	2.28	16.1	2.52	α	α	α	α	0.24	1.00	0.06	0.95
	63	12.0	0.58	20.9	1.13	19.7	0.73	30.8	0.34	16.5	2.00	24.1	2.18	α	α	α	α	0.04	0.94	0.11	1.00
	90	13.0	0.49	32.0	1.14	21.3	0.83	37.4	0.46	18.6	2.36	31.3	1.72	α	α	0.074	α	0.04	0.92	0.63	0.97
Lev	37	11.6	1.79	7.8	0.73	17.3	1.59	19.5	2.53	12.5	2.28	16.1	2.52	0.09	α	α	α	0.37	0.96	0.00	0.85
	63	14.2	1.76	17.3	3.28	19.8	1.17	30.6	2.99	16.5	2.00	24.1	2.18	α	α	α	α	0.19	0.93	0.05	0.99
	90	16.3	1.12	30.8	1.35	21.0	0.84	34.9	0.1	18.6	2.36	31.3	1.72	α	α	0.231	α	0.16	0.89	0.41	1.00
Man	37	11.4	0.93	8.8	2.03	18.0	0.57	21.6	1.08	12.5	2.28	16.1	2.52	0.09	α	α	α	0.32	1.00	0.02	0.97
	63	13.0	0.97	21.7	1.85	19.7	0.65	31.1	1.52	16.5	2.00	24.1	2.18	α	α	α	α	0.06	0.94	0.19	1.00
	90	14.1	0.51	32.0	1.83	20.9	0.77	37.2	4.66	18.6	2.36	31.3	1.72	α	α	0.193	α	0.05	0.88	0.60	0.96
NCD	37	7.5	2.42	10.2	2.02	17.6	1.41	19.5	1.41	12.5	2.28	16.1	2.52	0.09	α	α	α	0.08	0.97	0.03	0.87
	63	10.9	2.51	20.8	1.77	20.0	1.02	26.7	1.01	16.5	2.00	24.1	2.18	α	α	α	α	0.03	0.95	0.12	0.86
	90	15.0	2.24	31.9	1.11	21.3	1.08	34.5	1.35	18.6	2.36	31.3	1.72	α	α	0.108	α	0.11	0.90	0.62	0.94

TABLE II: The reduction analysis for the Cli project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	5	25.0	0.00	30.2	15.90	26.9	2.86	51.5	6.79	23.1	7.24	32.9	14.43	0.65	0.24	0.42	α	0.14	0.39	0.41	0.84
	8	25.0	0.00	40.4	14.32	29.8	2.64	60.6	2.86	24.2	7.35	49.2	10.17	0.18	α	0.01	α	0.14	0.69	0.30	0.86
	12	25.0	0.00	59.6	5.29	31.0	1.12	62.5	0.00	27.3	5.93	58.3	5.43	α	α	0.30	α	0.14	0.69	0.30	0.86
Ham	5	21.9	5.53	26.9	14.08	30.2	2.33	54.6	3.20	23.1	7.24	32.9	14.43	0.03	α	0.11	α	0.11	0.78	0.35	0.92
	8	22.1	5.53	37.5	8.54	31.3	0.00	59.0	3.10	24.2	7.35	49.2	10.17	α	α	α	α	0.11	0.81	0.18	0.79
	12	24.2	2.12	58.5	6.14	31.3	0.00	62.5	0.00	27.3	5.93	58.3	5.43	α	α	0.72	α	0.00	0.56	0.46	0.58
Lev	5	18.1	5.19	22.7	10.76	28.1	3.12	53.5	6.39	23.1	7.24	32.9	14.43	α	0.02	α	α	0.06	0.52	0.26	0.88
	8	21.3	4.45	40.8	12.15	30.0	2.50	61.0	2.64	24.2	7.35	49.2	10.17	α	α	α	α	0.09	0.70	0.27	0.87
	12	23.5	2.64	60.0	5.00	31.0	1.12	62.5	0.00	27.3	5.93	58.3	5.43	α	α	0.19	α	0.00	0.55	0.50	0.58
Man	5	25.0	0.00	28.1	14.77	28.1	3.12	50.0	6.04	23.1	7.24	32.9	14.43	0.65	0.02	0.18	α	0.14	0.52	0.37	0.80
	8	25.0	0.00	39.4	14.08	30.0	2.50	61.3	2.50	24.2	7.35	49.2	10.17	0.18	α	α	α	0.14	0.69	0.28	0.90
	12	25.0	0.00	60.6	4.88	30.6	1.88	62.5	0.00	27.3	5.93	58.3	5.43	α	0.02	0.09	α	0.00	0.52	0.53	0.58
NCD	5	24.4	1.88	25.8	10.42	25.8	5.53	46.7	8.50	23.1	7.24	32.9	14.43	0.34	0.63	0.07	α	0.13	0.45	0.33	0.76
	8	25.0	0.00	44.4	14.19	27.3	4.41	56.5	5.22	24.2	7.35	49.2	10.17	0.18	0.41	0.34	α	0.14	0.48	0.41	0.71
	12	25.0	0.00	50.8	7.86	31.0	1.12	62.5	0.00	27.3	5.93	58.3	5.43	α	α	α	α	0.00	0.55	0.20	0.58

TABLE III: The reduction analysis for the Codec project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	12	41.0	4.99	17.8	1.67	48.8	2.42	17.5	3.77	43.3	4.01	18.5	2.89	0.10	α	0.15	0.48	0.38	0.86	0.39	0.45
	22	46.6	2.70	25.2	1.47	52.8	2.07	25.9	1.55	46.3	3.34	24.8	2.46	0.73	α	0.59	0.14	0.52	0.97	0.54	0.61
	31	47.8	2.55	30.2	0.68	54.4	2.01	29.9	0.36	47.9	3.48	29.1	1.9	0.83	α	0.03	0.35	0.48	0.94	0.66	0.57
Ham	12	2.1	0.07	16.3	2.22	51.3	1.46	19.5	1.09	43.3	4.01	18.5	2.89	α	α	α	0.21	0.00	0.98	0.24	0.59
	22	2.6	2.47	22.5	1.98	52.6	1.74	26.4	1.56	46.3	3.34	24.8	2.46	α	α	α	0.02	0.00	0.97	0.26	0.68
	31	4.2	5.36	28.5	1.67	54.9	1.78	30.4	0.35	47.9	3.48	29.1	1.9	α	α	0.07	α	0.00	0.96	0.37	0.72
Lev	12	31.5	3.66	15.9	2.32	53.3	2.25	20.1	2.32	43.3	4.01	18.5	2.89	α	α	α	0.03	0.01	0.99	0.21	0.66
	22	38.9	3.28	24.8	2.40	55.5	0.47	27.6	0.61	46.3	3.34	24.8	2.46	α	α	0.93	α	0.06	1.00	0.50	0.86
	31	40.8	1.62	30.1	0.53	56.2	0.55	30.0	0.34	47.9	3.48	29.1	1.90	α	α	0.05	0.23	0.03	1.00	0.64	0.59
Man	12	37.0	5.00	17.7	1.20	50.3	2.38	19.3	1.63	43.3	4.01	18.5	2.89	α	α	0.05	0.34	0.19	0.92	0.36	0.57
	22	45.3	2.57	25.1	1.47	54.1	1.98	26.0	1.38	46.3	3.34	24.8	2.46	0.12	α	0.75	0.07	0.38	0.99	0.51	0.63
	31	47.5	2.26	30.2	0.55	55.6	0.82	30.3	0.30	47.9	3.48	29.1	1.90	0.48	α	0.03	α	0.44	0.98	0.67	0.69
NCD	12	37.0	4.16	17.7	0.68	50.3	2.14	19.3	1.62	43.3	4.01	18.5	2.89	α	α	0.91	α	0.00	0.94	0.49	0.77
	22	45.3	2.80	25.1	2.42	54.1	2.11	26.0	1.12	46.3	3.34	24.8	2.46	α	α	0.11	α	0.00	0.97	0.62	0.80
	31	47.5	0.54	30.2	1.27	55.6	1.87	30.3	0.27	47.9	3.48	29.1	1.90	α	α	0.27	α	0.00	0.90	0.58	0.81

TABLE IV: The reduction analysis for the JacksonCore project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	18	0.9	0.95	33.4	0.52	26.7	1.55	42.4	1.44	11.4	3.23	35.2	3.89	α	α	0.01	α	0.00	1.00	0.31	0.99
	31	2.1	0.74	38.0	1.03	29.7	1.29	51.4	1.26	15.9	3.21	45.3	4.23	α	α	α	α	0.00	1.00	0.07	0.95
	45	2.7	0.94	46.8	9.84	33.0	1.13	55.6	0.42	18.4	4.33	53.6	3.88	α	α	α	α	0.00	1.00	0.12	0.92
Ham	18	0.7	0.65	29.2	5.75	24.9	2.55	44.5	1.40	11.4	3.23	35.2	3.89	α	α	α	α	0.00	1.00	0.14	1.00
	31	1.0	0.60	37.9	3.74	29.7	1.64	52.8	0.96	15.9	3.21	45.3	4.23	α	α	α	α	0.00	1.00	0.09	0.99
	45	1.5	0.94	45.9	9.79	32.6	1.02	55.6	0.30	18.4	4.33	53.6	3.88	α	α	α	α	0.00	1.00	0.11	0.94
Lev	18	0.5	0.40	25.7	3.77	26.1	2.15	45.6	2.09	11.4	3.23	35.2	3.89	α	α	α	α	0.00	1.00	0.04	0.99
	31	0.6	0.41	37.7	2.13	29.4	1.33	50.4	9.39	15.9	3.21	45.3	4.23	α	α	α	α	0.00	1.00	0.11	0.94
	45	0.9	0.33	44.4	4.17	30.5	1.45	54.9	0.58	18.4	4.33	53.6	3.88	α	α	α	α	0.00	1.00	0.04	0.74
Man	18	0.2	0.00	32.7	1.69	26.5	1.36	43.3	1.46	11.4	3.23	35.2	3.89	α	α	α	α	0.00	1.00	0.28	0.99
	31	0.2	0.00	38.3	0.96	30.2	1.48	52.4	0.76	15.9	3.21	45.3	4.23	α	α	α	α	0.00	1.00	0.08	0.98
	45	0.6	0.71	47.8	5.12	32.8	1.01	55.2	0.50	18.4	4.33	53.6	3.88	α	α	α	α	0.00	1.00	0.14	0.84
NCD	18	0.2	0.00	31.3	2.42	23.4	2.92	43.7	4.63	11.4	3.23	35.2	3.89	α	α	α	α	0.00	1.00	0.20	0.97
	31	0.2	0.00	48.2	3.69	26.6	2.73	51.1	3.03	15.9	3.21	45.3	4.23	α	α	α	α	0.00	0.99	0.72	0.91
	45	0.2	0.00	54.8	0.37	29.2	2.30	54.6	0.76	18.4	4.33	53.6	3.88	α	α	α	0.01	0.00	0.99	0.73	0.67

TABLE V: The reduction analysis for the JxPath project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	19	9.6	2.40	34.0	2.54	11.6	1.48	31.8	4.30	9.1	3.31	26.6	5.68	0.77	α	α	α	0.52	0.69	0.88	0.75
	33	12.9	2.10	40.6	0.75	13.9	1.75	43.3	2.84	10.8	3.23	38.8	3.99	α	α	0.14	α	0.73	0.77	0.61	0.80
	47	14.4	1.25	47.1	0.48	16.0	1.74	52.2	0.67	12.9	2.69	50.8	2.78	0.05	α	α	0.17	0.65	0.81	0.14	0.59
Ham	19	6.0	1.87	33.1	4.73	11.6	1.22	34.0	4.87	9.1	3.31	26.6	5.68	α	0.02	α	α	0.14	0.68	0.80	0.83
	33	8.4	2.50	41.6	1.15	13.5	4.43	44.4	4.88	10.8	3.23	38.8	3.99	α	α	0.06	α	0.17	0.78	0.64	0.81
	47	10.3	2.33	46.5	0.64	14.8	0.57	50.1	2.45	12.9	2.69	50.8	2.78	α	0.01	α	0.39	0.24	0.69	0.11	0.44
Lev	19	3.3	1.01	22.6	3.70	11.5	1.71	35.4	3.33	9.1	3.31	26.6	5.68	α	0.01	α	α	0.01	0.69	0.29	0.88
	33	6.3	1.79	41.0	4.09	14.7	2.09	47.4	3.49	10.8	3.23	38.8	3.99	α	α	0.09	α	0.08	0.85	0.63	0.96
	47	8.2	1.68	51.8	9.12	16.5	1.71	53.6	1.25	12.9	2.69	50.8	2.78	α	α	0.16	α	0.07	0.86	0.39	0.85
Man	19	9.3	2.47	32.3	3.88	11.9	1.43	30.8	5.92	9.1	3.31	26.6	5.68	0.82	α	α	α	0.48	0.72	0.79	0.69
	33	12.5	2.16	42.8	4.74	14.5	1.46	45.0	4.14	10.8	3.23	38.8	3.99	0.03	α	0.01	α	0.68	0.87	0.69	0.85
	47	14.7	0.33	50.2	4.89	16.5	1.49	52.7	0.70	12.9	2.69	50.8	2.78	0.02	α	0.12	α	0.67	0.87	0.37	0.72
NCD	19	7.5	2.74	21.2	3.45	13.5	1.68	35.4	3.02	9.1	3.31	26.6	5.68	0.01	α	α	α	0.30	0.87	0.22	0.92
	33	10.6	2.19	36.5	2.20	16.2	1.60	46.2	6.86	10.8	3.23	38.8	3.99	0.14	α	0.02	α	0.38	0.94	0.31	0.88
	47	12.9	0.77	45.7	1.47	18.3	1.42	53.7	1.10	12.9	2.69	50.8	2.78	0.86	α	α	α	0.48	0.96	0.08	0.83

TABLE VI: The reduction analysis for the Math project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	39	0.0	0.00	30.9	2.74	26.5	0.97	38.0	3.17	16.2	3.60	37.1	3.88	α	α	α	0.40	0.00	1.00	0.11	0.56
	67	0.6	2.02	45.3	1.34	29.2	1.14	52.0	1.62	21.0	4.91	49.8	6.20	α	α	α	α	0.00	0.93	0.15	0.76
	95	2.2	3.03	54.5	1.25	31.5	1.69	60.9	0.52	25.3	3.70	56.0	2.89	α	α	α	α	0.00	0.89	0.05	0.90
Ham	39	6.2	3.02	29.9	2.47	26.8	0.16	39.3	1.67	16.2	3.60	37.1	3.88	α	α	α	0.03	0.01	1.00	0.07	0.66
	67	8.3	2.24	44.6	1.34	28.3	0.20	51.8	1.73	21.0	4.91	49.8	6.20	α	α	α	α	0.00	0.93	0.13	0.76
	95	9.3	2.60	56.0	2.64	31.7	1.24	58.6	1.18	25.3	3.70	59.0	2.89	α	α	α	0.23	0.00	0.92	0.10	0.40
Lev	39	0.0	0.00	30.0	2.97	26.6	0.95	40.0	1.61	16.2	3.60	37.1	3.88	α	α	α	α	0.00	1.00	0.09	0.74
	67	0.0	0.00	44.2	1.95	31.4	0.93	51.7	1.53	21.0	4.91	49.8	6.20	α	α	α	α	0.00	0.93	0.12	0.76
	95	0.1	0.65	55.7	1.40	33.8	0.34	59.3	1.48	25.3	3.70	59.0	2.89	α	α	α	0.15	0.00	0.99	0.08	0.61
Man	39	0.0	0.00	29.4	2.29	26.5	1.37	40.2	2.32	16.2	3.60	37.1	3.88	α	α	α	α	0.00	1.00	0.07	0.73
	67	0.84	2.30	43.7	0.99	29.0	0.84	51.5	2.13	21.0	4.91	49.8	6.20	α	α	α	α	0.00	0.93	0.10	0.72
	95	2.5	4.57	54.6	1.39	31.2	0.94	61.3	0.95	25.3	3.70	59.0	2.89	α	α	α	α	0.00	0.89	0.06	0.91
NCD	39	0.0	0.00	36.4	5.73	28.8	2.10	40.0	2.20	16.2	3.60	37.1	3.88	α	α	α	α	0.00	1.00	0.07	0.73
	67	0.00	0.00	51.6	1.24	31.9	1.30	50.1	1.52	21.0	4.91	49.8	6.20	α	α	α	α	0.00	0.93	0.10	0.72
	95	0.0	0.00	57.6	0.76	34.3	1.50	59.1	0.32	25.3	3.70	59.0	2.89	α	α	α	α	0.00	0.89	0.06	0.91

TABLE VII: The reduction analysis for the Compress project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	7	1.4	0.93	35.8	1.20	5.1	0.85	26.6	11.21	4.4	1.63	31.9	10.82	α	0.23	0.44	0.07	0.05	0.59	0.44	0.36
	13	2.6	1.41	41.9	2.01	6.1	0.76	44.9	3.68	6.5	1.41	41.4	8.52	α	0.03	0.17	0.18	0.04	0.34	0.40	0.60
	18	3.2	1.18	47.2	2.24	6.6	0.84	48.6	3.05	6.7	1.63	48.4	10.82	α	0.39	0.01	0.32	0.03	0.42	0.31	0.57
Ham	7	0.7	0.75	36.7	7.54	4.3	0.57	36.3	1.19	4.4	1.63	31.9	10.82	α	0.43	0.05	0.96	0.01	0.43	0.64	0.50
	13	2.2	1.30	44.4	3.11	4.7	0.43	40.9	1.93	6.5	1.41	41.4	8.52	α	α	0.38	0.02	0.02	0.10	0.56	0.33
	18	2.9	1.34	47.7	2.57	4.9	0.64	51.7	0.21	6.7	1.20	48.4	3.06	α	α	0.19	α	0.03	0.10	0.39	0.96
Lev	7	2.0	0.68	16.6	6.01	5.0	0.56	41.1	2.56	4.4	1.63	31.9	10.82	α	0.28	α	α	0.08	0.57	0.15	0.79
	13	3.0	0.98	36.3	12.16	6.1	0.70	48.7	2.36	6.5	1.41	41.4	8.52	α	0.01	0.27	α	0.04	0.30	0.41	0.89
	18	3.5	0.90	48.9	3.49	6.8	0.62	51.7	0.16	6.7	1.20	48.4	3.06	α	0.64	0.07	α	0.04	0.45	0.63	0.98
Man	7	1.8	0.98	34.2	6.16	4.7	0.62	28.54	12.05	4.4	1.63	31.9	10.82	α	0.73	0.29	0.23	0.08	0.51	0.42	0.41
	13	2.7	1.20	42.4	1.64	5.4	0.67	44.7	3.65	6.5	1.41	41.4	8.52	α	α	0.28	0.17	0.03	0.19	0.42	0.60
	18	3.0	1.53	47.8	2.37	5.6	0.71	48.7	3.02	6.7	1.20	48.4	3.06	α	α	0.11	0.23	0.03	0.18	0.37	0.58
NCD	7	1.4	0.41	13.1	4.48	4.4	0.91	42.8	2.34	4.4	1.63	31.9	10.82	α	0.71	α	α	0.02	0.47	0.06	0.87
	13	2.1	0.58	25.2	11.92	5.4	0.57	49.3	0.42	6.5	1.41	41.4	8.52	α	α	α	α	0.00	0.17	0.13	0.92
	18	2.4	0.32	42.8	5.55	6.0	0.84	51.7	0.21	6.7	1.20	48.4	3.06	α	α	α	α	0.00	0.26	0.12	0.97

TABLE VIII: The reduction analysis for the Csv project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	7	2.2	2.08	16.8	2.98	2.0	2.09	26.2	3.24	2.4	2.19	18.2	6.87	0.40	0.24	0.55	α	0.24	0.22	0.45	0.84
	13	1.8	1.57	33.1	5.23	1.5	1.69	41.3	3.04	3.6	2.91	33.4	6.46	α	α	0.72	α	0.17	0.09	0.52	0.84
	18	3.1	2.42	46.5	4.76	2.0	2.32	49.0	0.57	5.1	3.67	45.2	3.75	0.04	α	0.18	α	0.32	0.12	0.58	0.83
Ham	7	4.6	1.27	16.0	3.15	1.0	0.00	28.1	3.26	2.4	2.19	18.2	6.87	α	α	0.26	α	0.83	0.00	0.41	0.88
	13	5.0	0.73	30.5	5.55	1.0	0.00	38.5	3.50	3.6	2.91	33.4	6.46	α	α	0.08	α	0.80	0.00	0.37	0.75
	18	5.3	0.88	48.8	4.60	1.0	0.00	49.1	0.56	5.1	3.67	45.2	3.75	0.02	α	α	α	0.65	0.00	0.75	0.84
Lev	7	2.1	1.75	14.4	4.32	1.0	1.50	29.3	2.89	2.4	2.19	18.2	6.87	0.49	0.02	0.04	α	0.27	0.10	0.34	0.91
	13	4.0	2.02	23.7	3.10	1.0	1.72	39.6	2.09	3.6	2.91	33.4	6.46	0.15	α	α	α	0.57	0.09	0.07	0.81
	18	4.1	2.14	40.1	2.98	1.4	1.51	48.0	1.16	5.1	3.67	45.2	3.75	0.97	α	α	α	0.48	0.05	0.16	0.73
Man	7	2.4	1.98	17.3	2.20	1.5	1.16	27.1	3.08	2.4	2.19	18.2	6.87	0.87	0.05	0.77	α	0.32	0.13	0.48	0.87
	13	2.9	2.35	36.1	3.72	1.3	0.95	39.9	3.51	3.6	2.91	33.4	6.46	0.27	α	α	α	0.32	0.07	0.67	0.82
	18	3.1	2.16	49.3	1.97	1.4	0.96	49.0	0.50	5.1	3.67	45.2	3.75	0.06	α	α	α	0.34	0.04	0.84	0.83
NCD	7	2.0	1.38	11.3	2.64	2.0	3.06	23.3	4.53	2.4	2.19	18.2	6.87	0.60	0.07	α	α	0.28	0.14	0.18	0.73
	13	3.1	1.83	27.4	2.53	1.3	0.56	44.0	2.50	3.6	2.91	33.4	6.46	0.95	α	α	α	0.44	0.07	0.20	0.90
	18	4.4	1.15	44.8	2.98	1.5	1.22	49.1	0.49	5.1	3.67	45.2	3.75	0.33	α	0.51	α	0.55	0.08	0.44	0.85

TABLE IX: The reduction analysis for the Gson project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	23	9.7	0.84	16.8	1.17	13.1	1.68	25.4	3.53	11.4	1.73	25.5	4.48	α	α	α	0.77	0.16	0.81	0.04	0.47
	39	10.7	1.27	23.9	1.65	14.3	2.11	33.9	2.80	12.3	1.47	35.2	3.81	α	α	α	0.09	0.18	0.77	0.00	0.36
	56	12.6	1.37	39.1	3.44	14.3	1.83	46.2	1.67	13.4	1.42	44.0	2.88	0.04	0.09	α	α	0.33	0.60	0.13	0.72
Ham	23	6.5	0.39	19.1	3.99	13.0	0.25	23.9	1.00	11.4	1.73	25.5	4.48	α	α	α	0.21	0.00	0.84	0.15	0.40
	39	10.0	0.51	26.5	3.21	13.1	0.33	30.7	1.18	12.3	1.47	35.2	3.81	α	0.02	α	α	0.00	0.68	0.04	0.18
	56	8.0	1.21	44.1	3.54	13.4	0.24	45.7	4.50	13.4	1.42	44.0	2.88	α	0.91	0.36	0.52	0.00	0.50	0.56	0.54
Lev	23	5.8	0.73	22.0	3.48	14.1	1.63	25.1	3.36	11.4	1.73	25.5	4.48	α	α	α	0.62	0.00	0.89	0.30	0.46
	39	6.7	0.27	29.1	2.58	14.2	0.47	36.3	1.52	12.3	1.47	35.2	3.81	α	α	α	0.71	0.00	0.87	0.13	0.53
	56	6.9	0.81	42.3	3.73	9.0	1.66	43.4	8.08	13.4	1.42	44.0	2.88	α	α	0.04	0.97	0.00	0.79	0.35	0.49
Man	23	9.6	0.58	16.7	0.97	13.6	1.98	26.1	2.61	11.4	1.73	25.5	4.48	α	α	α	0.53	0.11	0.85	0.03	0.54
	39	10.6	0.96	23.2	1.67	13.9	1.91	34.1	3.78	12.3	1.47	35.2	3.81	α	α	α	0.34	0.16	0.73	0.00	0.42
	56	12.1	1.32	41.4	3.48	13.6	1.12	45.3	1.12	13.4	1.42	44.0	2.88	α	0.81	α	0.19	0.23	0.51	0.25	0.59
NCD	23	5.3	0.28	15.9	1.77	14.9	2.56	24.1	3.18	11.4	1.73	25.5	4.48	α	α	α	0.17	0.00	0.89	0.03	0.39
	39	6.6	0.30	25.9	3.31	15.6	2.14	36.9	1.03	12.3	1.47	35.2	3.81	α	α	α	0.44	0.00	0.93	0.05	0.55
	56	6.8	0.20	42.0	4.15	15.8	2.13	43.1	1.16	13.4	1.42	44.0	2.88	α	α	0.06	α	0.00	0.84	0.35	0.30

TABLE X: The reduction analysis for the JacksonDatabind project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	44	16.5	2.32	35.3	0.78	18.2	0.21	42.0	0.94	15.1	2.26	39.7	1.65	0.24	α	α	α	0.58	0.80	0.02	0.80
	76	19.4	1.08	44.6	0.65	19.1	0.57	48.4	0.56	18.7	3.23	46.8	3.89	0.51	0.67	α	α	0.55	0.46	0.09	0.82
	108	20.5	0.39	49.6	0.57	20.1	1.05	52.0	0.20	20.2	1.36	50.9	0.93	0.11	0.52	α	α	0.60	0.43	0.12	0.90
Ham	44	16.8	2.64	34.5	1.02	18.2	0.17	42.6	0.79	15.1	2.26	39.7	1.65	0.05	α	α	α	0.64	0.80	0.01	0.90
	76	19.8	0.83	44.7	0.71	19.0	0.63	47.7	0.16	18.7	3.23	46.8	3.89	0.09	0.57	α	α	0.62	0.46	0.10	0.71
	108	20.5	0.23	49.6	0.50	19.3	0.48	52.3	0.13	20.2	1.36	50.9	0.93	0.04	α	α	α	0.65	0.19	0.12	0.95
Lev	44	13.0	4.45	35.9	1.21	18.8	0.77	42.0	0.75	15.1	2.26	39.7	1.65	0.02	α	α	α	0.33	0.85	0.09	0.81
	76	19.6	0.94	44.8	0.79	19.3	0.48	47.4	0.80	18.7	3.23	46.8	3.89	0.26	0.98	α	0.14	0.58	0.50	0.11	0.61
	108	20.3	0.17	49.6	0.54	21.6	1.19	52.1	0.23	20.2	1.36	50.9	0.93	0.27	α	α	α	0.57	0.76	0.13	0.92
Man	44	16.6	1.24	35.2	0.87	18.2	0.19	42.1	0.93	15.1	2.26	39.7	1.65	0.28	α	α	α	0.57	0.80	0.03	0.83
	76	19.3	1.17	44.5	0.54	19.4	0.88	47.9	0.43	18.7	3.23	46.8	3.89	0.58	0.74	α	α	0.54	0.52	0.08	0.73
	108	20.3	0.31	49.6	0.51	19.9	1.00	52.2	0.14	20.2	1.36	50.9	0.93	0.28	0.09	α	α	0.56	0.34	0.12	0.95
NCD	44	16.8	1.38	34.8	1.76	17.9	1.88	38.8	1.97	15.1	2.26	39.7	1.65	0.17	α	α	0.10	0.60	0.78	0.05	0.38
	76	19.3	0.99	45.2	1.31	19.6	0.83	46.8	1.33	18.7	3.23	46.8	3.89	0.55	0.36	α	0.89	0.54	0.56	0.21	0.49
	108	20.3	0.06	50.3	0.45	20.8	0.97	51.3	0.59	20.2	1.36	50.9	0.93	0.30	0.02	α	0.10	0.56	0.66	0.26	0.61

TABLE XI: The reduction analysis for the Jsoup project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	6	15.3	1.44	15.5	0.33	21.4	7.48	14.6	1.36	19.4	9.02	14.2	1.39	0.19	0.16	α	0.37	0.39	0.60	0.81	0.55
	11	16.4	2.70	16.5	0.67	24.1	8.97	16.0	0.61	20.4	7.19	16.2	1.04	0.10	0.05	0.52	0.32	0.37	0.64	0.50	0.36
	16	17.6	3.10	17.3	0.15	28.5	7.33	16.6	0.74	25.5	7.86	17.0	0.87	α	0.06	0.63	α	0.22	0.63	0.44	0.27
Ham	6	5.1	3.16	14.3	1.07	16.9	4.18	14.3	1.06	19.4	9.02	14.2	1.39	α	0.64	0.98	0.25	0.03	0.45	0.49	0.56
	11	7.8	3.91	15.9	0.50	32.5	0.20	16.1	0.66	20.4	7.19	16.2	1.04	α	α	0.07	0.51	0.01	0.91	0.26	0.38
	16	10.4	3.54	16.4	0.68	32.5	0.09	16.9	0.79	25.5	7.86	17.0	0.87	α	α	α	0.28	0.00	0.78	0.21	0.40
Lev	6	5.8	3.89	15.1	0.68	24.7	8.99	14.3	1.29	19.4	9.02	14.2	1.39	α	0.23	α	0.51	0.05	0.67	0.72	0.52
	11	6.3	3.87	15.9	0.38	33.6	3.82	16.0	0.75	20.4	7.19	16.2	1.04	α	α	0.09	0.23	0.01	0.93	0.28	0.35
	16	10.7	3.23	16.3	0.50	35.1	1.72	16.4	3.15	25.5	7.86	17.0	0.87	α	α	α	0.60	0.00	0.89	0.19	0.44
Man	6	12.5	4.33	15.6	0.67	18.3	5.03	14.9	1.14	19.4	9.02	14.2	1.39	α	0.73	α	0.02	0.27	0.52	0.81	0.67
	11	15.9	1.47	16.2	0.64	23.2	7.68	15.9	0.88	20.4	7.19	16.2	1.04	0.12	0.10	0.71	0.22	0.38	0.61	0.42	0.33
	16	16.6	3.09	17.3	0.15	34.5	2.47	17.4	4.06	25.5	7.86	17.0	0.87	α	α	0.63	0.11	0.15	0.85	0.44	0.32
NCD	6	12.1	2.96	14.7	0.65	17.0	4.77	14.3	1.11	19.4	9.02	14.2	1.39	α	0.52	0.08	0.68	0.14	0.45	0.57	0.51
	11	15.3	5.93	15.8	0.25	26.4	8.51	15.8	0.97	20.4	7.19	16.2	1.04	α	α	0.05	0.08	0.19	0.71	0.27	0.32
	16	21.0	7.54	16.3	0.54	29.9	6.97	17.1	2.43	25.5	7.86	17.0	0.87	α	0.01	α	0.15	0.28	0.69	0.18	0.34

TABLE XII: The reduction analysis for the Lang project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	43	2.1	1.13	18.8	0.60	8.7	0.73	19.6	1.03	7.9	1.74	24.5	2.24	α	0.12	α	α	0.00	0.62	0.00	0.00
	73	3.6	1.17	32.4	1.75	9.9	0.67	29.6	1.29	11.4	1.90	34.7	2.73	α	α	α	α	0.00	0.21	0.26	0.03
	104	5.2	0.82	44.6	1.07	10.8	0.56	44.2	0.85	13.8	1.99	55.1	1.48	α	α	0.15	0.79	0.00	0.06	0.60	0.48
Ham	43	4.0	1.47	21.1	2.42	8.9	0.73	19.4	1.09	7.9	1.74	24.5	2.24	α	0.05	α	α	0.05	0.65	0.15	0.01
	73	5.5	1.54	33.8	2.03	11.8	0.86	28.3	0.63	11.4	1.90	34.7	2.73	α	0.70	0.34	α	0.01	0.53	0.43	0.00
	104	6.4	1.77	45.8	0.59	12.6	0.92	42.9	1.74	13.8	1.99	55.1	1.48	α	α	α	0.01	0.00	0.28	0.88	0.31
Lev	43	1.3	0.12	25.4	1.37	9.5	1.18	19.7	1.73	7.9	1.74	24.5	2.24	α	α	0.16	α	0.00	0.76	0.60	0.04
	73	2.0	0.23	35.4	0.76	11.1	0.86	27.6	1.58	11.4	1.90	34.7	2.73	α	0.16	0.13	α	0.00	0.39	0.62	0.00
	104	2.4	0.12	45.5	0.38	12.3	1.06	43.1	0.64	13.8	1.99	55.1	1.48	α	α	α	α	0.00	0.25	0.82	0.26
Man	43	1.9	0.68	18.9	0.91	8.5	0.76	18.7	1.23	7.9	1.74	24.5	2.24	α	0.33	α	α	0.00	0.57	0.00	0.00
	73	3.0	1.13	32.8	1.88	10.3	0.77	29.0	1.16	11.4	1.90	34.7	2.73	α	α	α	α	0.00	0.26	0.28	0.01
	104	4.8	0.95	44.2	1.07	10.8	0.70	42.8	1.48	13.8	1.99	55.1	1.48	α	α	0.81	α	0.00	0.07	0.52	0.27
NCD	43	2.7	0.22	24.9	1.95	10.9	0.82	20.7	2.09	7.9	1.74	24.5	2.24	α	α	0.60	α	0.00	0.94	0.54	0.12
	73	3.6	0.43	36.1	0.70	13.5	1.45	28.9	1.22	11.4	1.90	34.7	2.73	α	α	α	α	0.00	0.81	0.70	0.00
	104	4.4	0.45	44.9	0.41	14.9	1.16	43.1	0.92	13.8	1.99	55.1	1.48	α	0.03	0.03	α	0.00	0.66	0.66	0.28

TABLE XIII: The reduction analysis for the Time project.

SM	TS	Least diverse				Most diverse				Random diverse				p-Value				Effect size			
		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo		Rand		Evo	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	LR	MR	LR	MR	LR	MR	LR	MR
Euc	26	7.2	0.74	28.9	4.67	8.5	0.42	35.9	4.71	8.3	0.71	36.5	9.92	α	0.24	α	0.07	0.12	0.58	0.15	0.36
	45	8.3	0.43	39.7	2.78	9.0	0.29	43.0	4.18	8.6	0.60	49.1	4.24	0.02	0.04	α	α	0.32	0.64	0.04	0.15
	63	8.7	0.22	58.1	0.56	8.9	0.27	53.4	1.45	9.1	0.43	58.9	1.64	α	α	α	α	0.18	0.26	0.27	0.02
Ham	26	2.5	1.06	31.3	3.18	8.6	0.11	30.5	2.17	8.3	0.71	36.5	9.92	α	0.03	α	α	0.00	0.66	0.19	0.18
	45	4.7	1.19	40.4	2.01	8.7	0.00	47.3	1.04	8.6	0.60	49.1	4.24	α	0.66	α	0.06	0.00	0.43	0.02	0.35
	63	5.8	0.96	54.7	2.55	8.7	0.09	54.5	6.02	9.1	0.43	58.9	1.64	α	α	α	α	0.00	0.17	0.09	0.12
Lev	26	2.5	1.41	36.6	3.39	8.0	0.31	36.2	5.23	8.3	0.71	36.5	9.92	α	0.02	0.19	0.19	0.00	0.30	0.40	0.40
	45	5.1	1.45	45.4	1.64	8.3	0.42	51.9	2.29	8.6	0.60	49.1	4.24	α	0.01	α	0.02	0.01	0.31	0.24	0.67
	63	5.9	1.30	57.5	0.79	8.6	0.46	61.0	0.43	9.1	0.43	58.9	1.64	α	α	α	α	0.00	0.21	0.18	0.93
Man	26	6.4	0.82	29.7	3.54	8.6	0.30	35.6	5.01	8.3	0.71	36.5	9.92	α	0.04	α	0.08	0.04	0.64	0.16	0.37
	45	7.6	0.97	40.7	3.75	8.7	0.12	47.9	1.95	8.6	0.60	49.1	4.24	α	0.95	α	0.11	0.17	0.48	0.08	0.38
	63	8.6	0.20	58.1	0.60	8.8	0.15	55.9	2.63	9.1	0.43	58.9	1.64	α	α	α	α	0.13	0.20	0.27	0.18
NCD	26	7.4	0.91	26.2	9.33	8.2	0.65	36.0	5.18	8.3	0.71	36.5	9.92	α	0.34	α	0.17	0.21	0.40	0.15	0.40
	45	8.1	0.44	49.2	2.51	8.6	0.57	53.0	2.80	8.6	0.60	49.1	4.24	α	0.67	0.87	α	0.22	0.45	0.48	0.74
	63	8.2	0.09	59.1	1.09	9.1	0.52	61.4	0.41	9.1	0.43	58.9	1.64	α	0.89	0.93	α	0.03	0.43	0.50	0.96