Supplementary Material: Robot Causal Discovery Aided by Human Interaction

Filip Edström, Thomas Hellström, Xavier de Luna

March 23, 2023

Contents

1 Tables

Table 1: Best finishing (lowest SHD-EM) versions from the PC and MMHC starting points for all DAGs. *N* is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and Diff is the difference in Relative SHD-Em from Finish to Start.

DAG	N	Version	Start	Finish	Diff
Survey	500	MMHC Assoc	0.726 ± 0.0117	0.000 ± 0.0000	-0.726
	500	MMHC HC	0.726 ± 0.0117	0.000 ± 0.0000	-0.726
	500	MMHC Rndm	0.726 ± 0.0117	0.000 ± 0.0000	-0.726
	500	PC Assoc	0.814 ± 0.0112	0.000 ± 0.0000	-0.814
	500	PC HC	0.814 ± 0.0112	0.000 ± 0.0000	-0.814
	1000	MMHC Assoc	0.594 ± 0.0096	0.000 ± 0.0000	-0.594
	1000	MMHC HC	0.594 ± 0.0096	0.000 ± 0.0000	-0.594
	1000	MMHC Rndm	0.594 ± 0.0096	0.000 ± 0.0000	-0.594
	1000	PC Assoc	0.740 ± 0.0122	0.000 ± 0.0000	-0.740
	1000	PC HC	0.740 ± 0.0122	0.000 ± 0.0000	-0.740
	1500	MMHC Assoc	0.541 ± 0.0095	0.000 ± 0.0000	-0.541
	1500	MMHC HC	0.541 ± 0.0095	0.000 ± 0.0000	-0.541
	1500	PC Assoc	0.691 ± 0.0129	0.000 ± 0.0000	-0.691
	1500	PC HC	0.691 ± 0.0129	0.000 ± 0.0000	-0.691
	3000	MMHC Assoc	0.459 ± 0.0096	0.000 ± 0.0000	-0.459
	3000	MMHC HC	0.459 ± 0.0096	0.000 ± 0.0000	-0.459
	3000	PC Assoc	0.572 ± 0.0152	0.000 ± 0.0000	-0.572
	3000	PC HC	0.572 ± 0.0152	0.000 ± 0.0000	-0.572
Asia	500	MMHC HC	0.461 ± 0.0069	0.001 ± 0.0010	-0.460
	500	PC HC	0.728 ± 0.0076	0.007 ± 0.0019	-0.722
	1000	MMHC HC	0.416 ± 0.0067	0.000 ± 0.0000	-0.416
	1000	PC HC	0.676 ± 0.0080	0.003 ± 0.0011	-0.674
	1500	MMHC HC	0.394 ± 0.0055	0.000 ± 0.0000	-0.394
	1500	PC HC	0.628 ± 0.0067	0.002 ± 0.0009	-0.626
	3000	MMHC HC	0.371 ± 0.0042	0.000 ± 0.0000	-0.371
	3000	PC HC	0.586 ± 0.0054	0.000 ± 0.0000	-0.58

Table 1: Best finishing (lowest SHD-EM) versions from the PC and MMHC starting points for all DAGs. *N* is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and Diff is the difference in Relative SHD-Em from Finish to Start.

DAG	N	Version	Start	Finish	Diff
Sachs	500	MMHC HC	0.854 ± 0.0045	0.155 ± 0.0034	-0.699
	500	PC HC	0.844 ± 0.0040	0.132 ± 0.0033	-0.713
	1000	MMHC HC	0.864 ± 0.0041	0.092 ± 0.0043	-0.772
	1000	PC HC	0.736 ± 0.0044	0.062 ± 0.0017	-0.674
	1500	MMHC HC	0.861 ± 0.0032	0.019 ± 0.0039	-0.841
	1500	PC HC	0.800 ± 0.0041	0.089 ± 0.0048	-0.710
	3000	MMHC HC	0.847 ± 0.0036	0.062 ± 0.0026	-0.785
	3000	PC HC	0.741 ± 0.0034	0.064 ± 0.0042	-0.677
Child	500	MMHC HC	0.603 ± 0.0041	0.214 ± 0.0040	-0.389
	500	PC HC	0.746 ± 0.0039	0.174 ± 0.0031	-0.572
	1000	MMHC HC	0.531 ± 0.0039	0.125 ± 0.0044	-0.406
	1000	PC HC	0.660 ± 0.0054	0.108 ± 0.0040	-0.552
	1500	MMHC HC	0.511 ± 0.0041	0.117 ± 0.0058	-0.394
	1500	PC HC	0.621 ± 0.0043	0.055 ± 0.0027	-0.565
	3000	MMHC HC	0.462 ± 0.0040	0.124 ± 0.0090	-0.337
	3000	PC HC	0.588 ± 0.0055	0.007 ± 0.0019	-0.581
Insurance	500	MMHC HC	0.806 ± 0.0025	0.605 ± 0.0029	-0.201
	500	PC HC	0.847 ± 0.0026	0.502 ± 0.0031	-0.345
	1000	MMHC HC	0.751 ± 0.0031	0.529 ± 0.0037	-0.221
	1000	PC HC	0.783 ± 0.0025	0.419 ± 0.0038	-0.364
	1500	MMHC HC	0.714 ± 0.0029	0.482 ± 0.0037	-0.231
	1500	PC HC	0.743 ± 0.0036	0.415 ± 0.0052	-0.328
	3000	MMHC HC	0.709 ± 0.0027	0.505 ± 0.0029	-0.204
	3000	PC HC	0.674 ± 0.0022	0.283 ± 0.0030	-0.391

Table 1: Best finishing (lowest SHD-EM) versions from the PC and MMHC starting points for all DAGs. N is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and Diff is the difference in Relative SHD-Em from Finish to Start.

DAG	N	Version	Start	Finish	Diff
Mildew	500	MMHC HC	0.911 ± 0.0006	0.717 ± 0.0012	-0.194
	500	PC Assoc	0.884 ± 0.0018	0.708 ± 0.0023	-0.176
	1000	MMHC HC	0.848 ± 0.0005	0.678 ± 0.0012	-0.169
	1000	PC HC	0.873 ± 0.0010	0.635 ± 0.0009	-0.238
	1500	MMHC HC	0.867 ± 0.0006	0.674 ± 0.0001	-0.193
	1500	PC HC	0.873 ± 0.0010	0.608 ± 0.0004	-0.265
	3000	MMHC HC	0.869 ± 0.0012	0.646 ± 0.0009	-0.222
	3000	PC HC	0.871 ± 0.0005	0.569 ± 0.0007	-0.302
	5000	MMHC HC	0.870 ± 0.0001	0.674 ± 0.0000	-0.196
	5000	PC HC	0.838 ± 0.0003	0.564 ± 0.0004	-0.274
Alarm	500	MMHC HC	0.687 ± 0.0034	0.434 ± 0.0055	-0.253
	500	PC HC	0.822 ± 0.0033	0.457 ± 0.0040	-0.366
	1000	MMHC HC	0.564 ± 0.0030	0.256 ± 0.0057	-0.308
	1000	PC HC	0.696 ± 0.0044	0.326 ± 0.0058	-0.371
	1500	MMHC HC	0.511 ± 0.0028	0.172 ± 0.0036	-0.339
	1500	PC HC	0.632 ± 0.0039	0.307 ± 0.0053	-0.325
	3000	MMHC HC	0.483 ± 0.0030	0.278 ± 0.0039	-0.204
	3000	PC HC	0.493 ± 0.0060	0.124 ± 0.0051	-0.369
	5000	MMHC HC	0.405 ± 0.0029	0.142 ± 0.0035	-0.263
	5000	PC HC	0.439 ± 0.0045	0.115 ± 0.0067	-0.325
Barley	500	MMHC HC	0.884 ± 0.0008	0.759 ± 0.0007	-0.125
	500	PC HC	0.985 ± 0.0007	0.770 ± 0.0008	-0.215
	1000	MMHC HC	0.854 ± 0.0006	0.694 ± 0.0006	-0.160
	1000	PC HC	0.972 ± 0.0009	0.771 ± 0.0008	-0.201
	1500	MMHC HC	0.844 ± 0.0007	0.691 ± 0.0007	-0.153
	1500	PC HC	0.961 ± 0.0014	0.767 ± 0.0010	-0.193
	3000	MMHC HC	0.781 ± 0.0012	0.576 ± 0.0017	-0.205
	3000	PC HC	0.861 ± 0.0016	0.666 ± 0.0017	-0.195
	5000	MMHC HC	0.735 ± 0.0011	0.509 ± 0.0018	-0.226
	5000	PC HC	0.863 ± 0.0010	0.665 ± 0.0010	-0.198

Table 1: Best finishing (lowest SHD-EM) versions from the PC and MMHC starting points for all DAGs. N is sample size, Start is starting Relative SHD-EM with 95% confidence bands, Finish is Relative SHD-EM after 30 questions with 95% confidence bands, and Diff is the difference in Relative SHD-Em from Finish to Start.

DAG	N	Version	Start	Finish	Diff
Hailfinder	500	MMHC HC	0.647 ± 0.0010	0.371 ± 0.0017	-0.276
	500	PC HC	0.713 ± 0.0019	0.415 ± 0.0022	-0.298
	1000	MMHC HC	0.601 ± 0.0012	0.296 ± 0.0016	-0.304
	1000	PC HC	0.652 ± 0.0020	0.325 ± 0.0020	-0.327
	1500	MMHC HC	0.565 ± 0.0012	0.248 ± 0.0015	-0.317
	1500	PC HC	0.627 ± 0.0026	0.297 ± 0.0025	-0.330
	3000	MMHC HC	0.541 ± 0.0013	0.195 ± 0.0016	-0.346
	3000	PC HC	0.605 ± 0.0028	0.260 ± 0.0029	-0.345
	5000	MMHC HC	0.525 ± 0.0010	0.150 ± 0.0013	-0.375
	5000	PC HC	0.605 ± 0.0029	0.250 ± 0.0032	-0.355

2 Figures

OptSingle sometimes doesn't ask any questions, e.g. Figure 1. This is due to not all 500 starting PDAGs having an undirected edge.

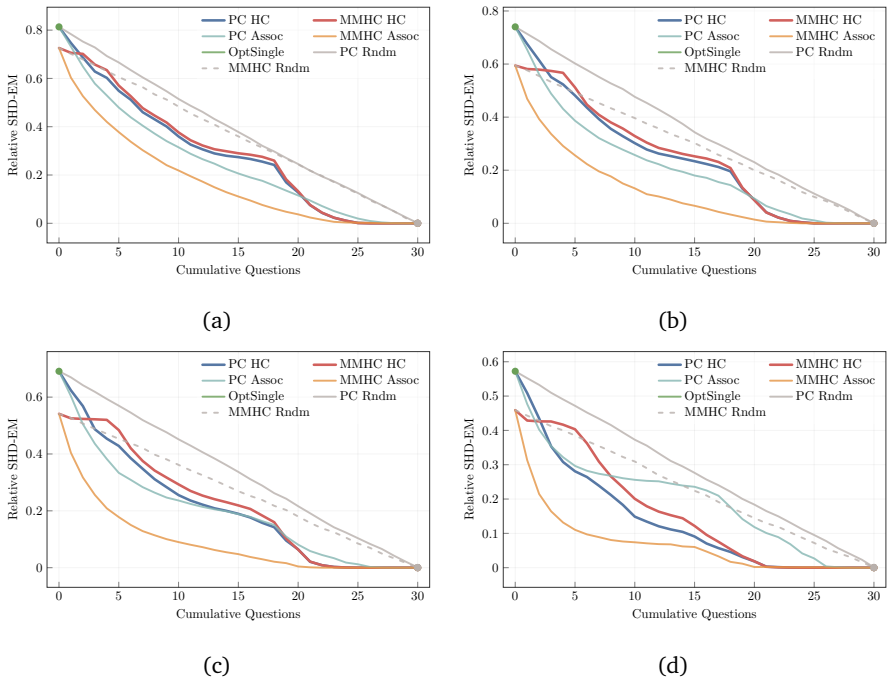

Figure 1: SHD-EM averaged over 500 replicates on the Survey DAG (6 nodes, 6 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$.

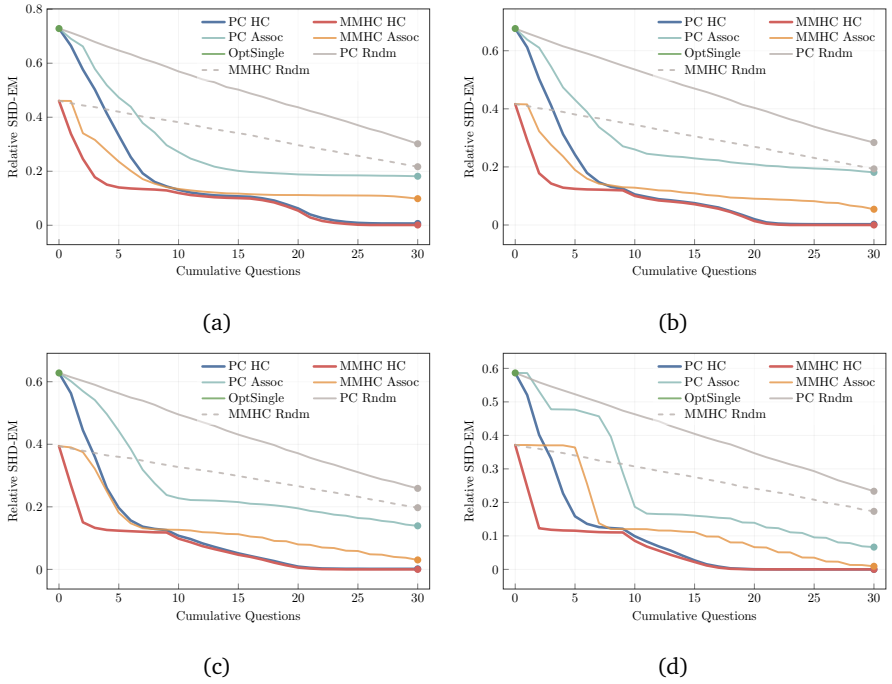

Figure 2: SHD-EM averaged over 500 replicates on the Asia DAG (8 nodes, 8 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$.

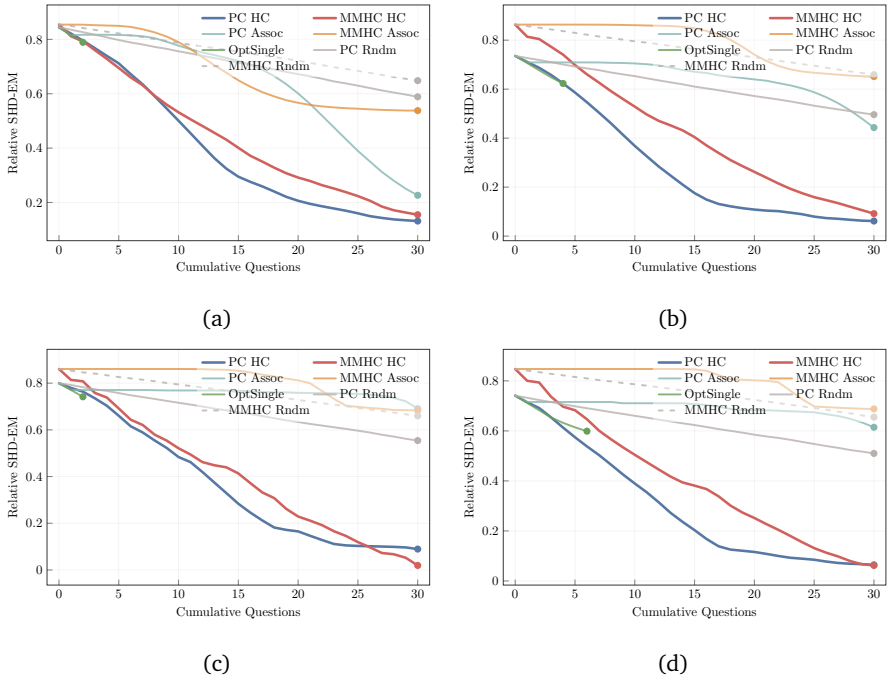

Figure 3: SHD-EM averaged over 500 replicates on the Sachs DAG (11 nodes, 17 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$.

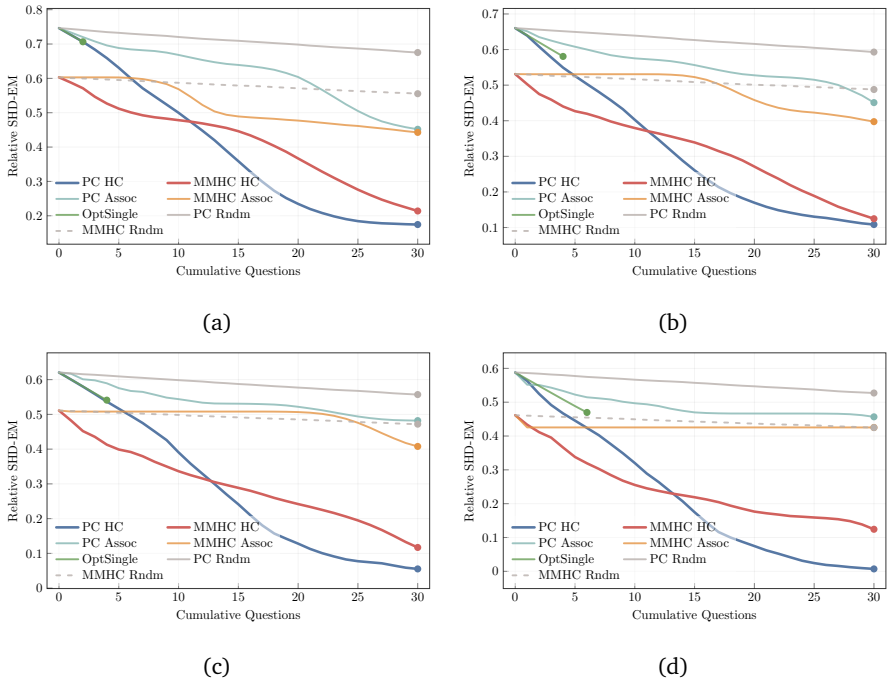

Figure 4: SHD-EM averaged over 500 replicates on the Child DAG (20 nodes, 25 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$.

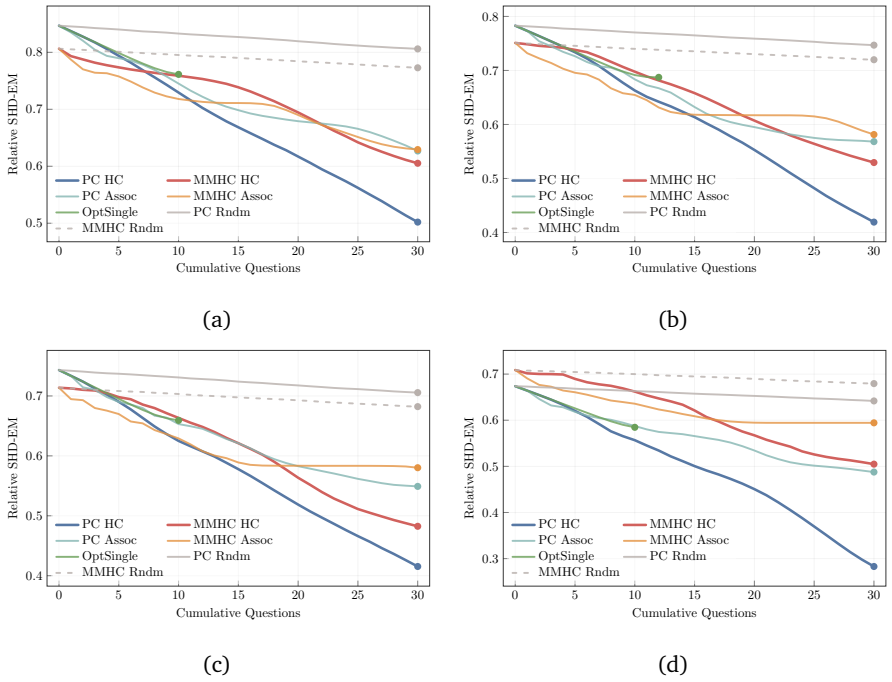

Figure 5: SHD-EM averaged over 500 replicates on the Insurance DAG (27 nodes, 52 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$.

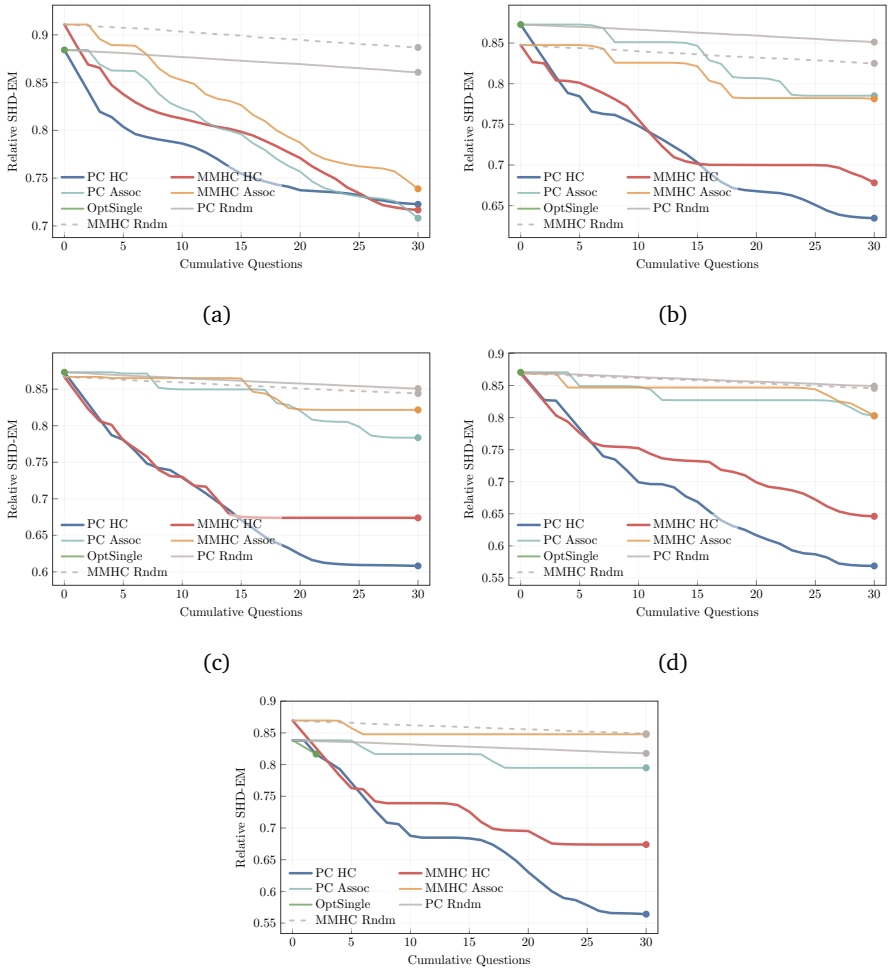

Figure 6: SHD-EM averaged over 500 replicates on the Mildew DAG (35 nodes, 46 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$, (e) $N = 5000$.

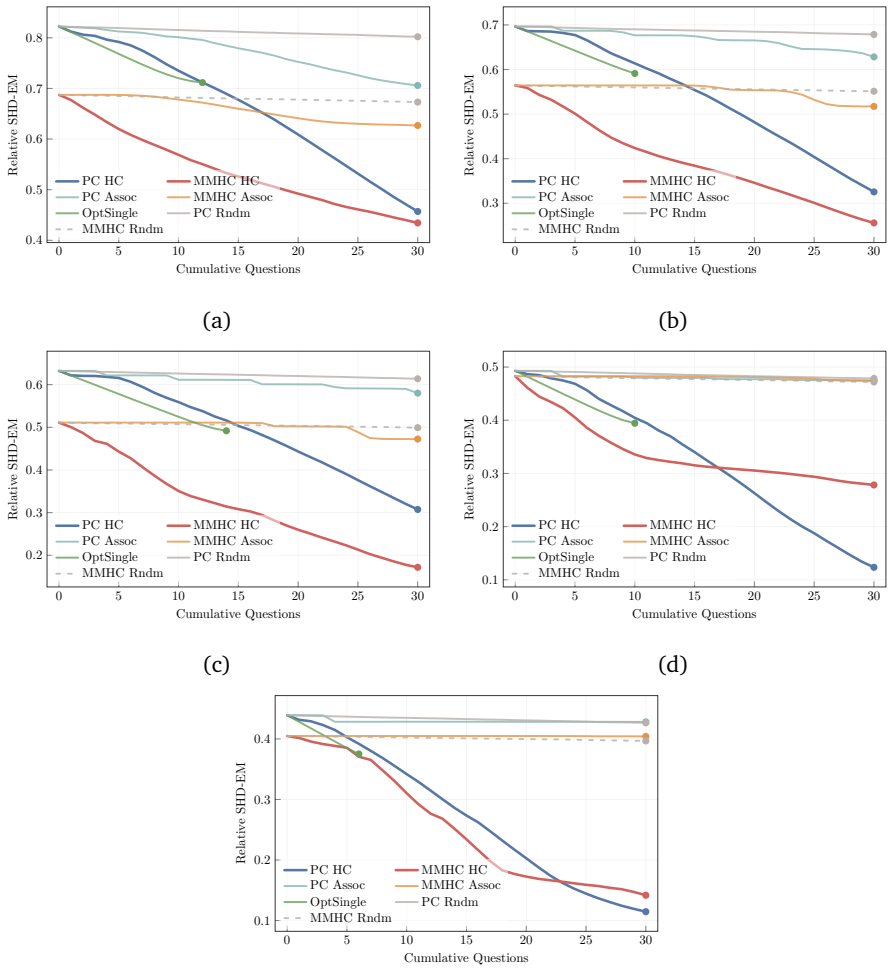

Figure 7: SHD-EM averaged over 500 replicates on the Alarm DAG, (37 nodes, 46 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$, (e) $N = 5000$.

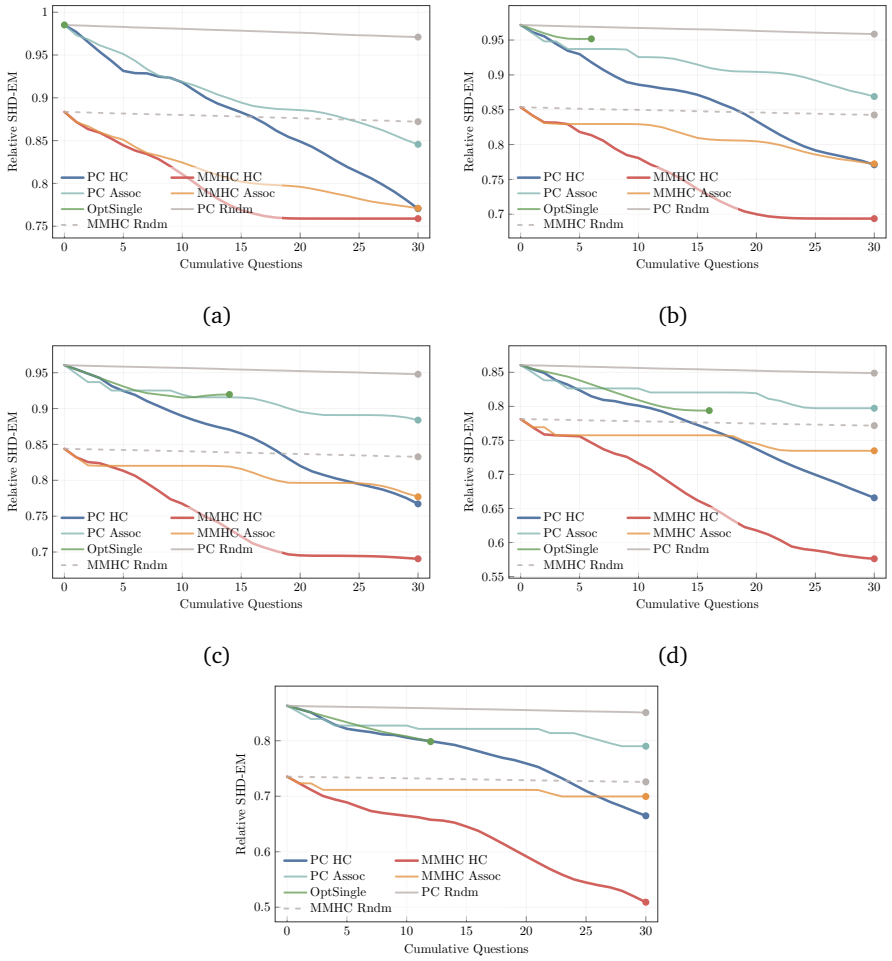

Figure 8: SHD-EM averaged over 500 replicates on the Barley DAG (48 nodes, 84 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$, (e) $N = 5000$.

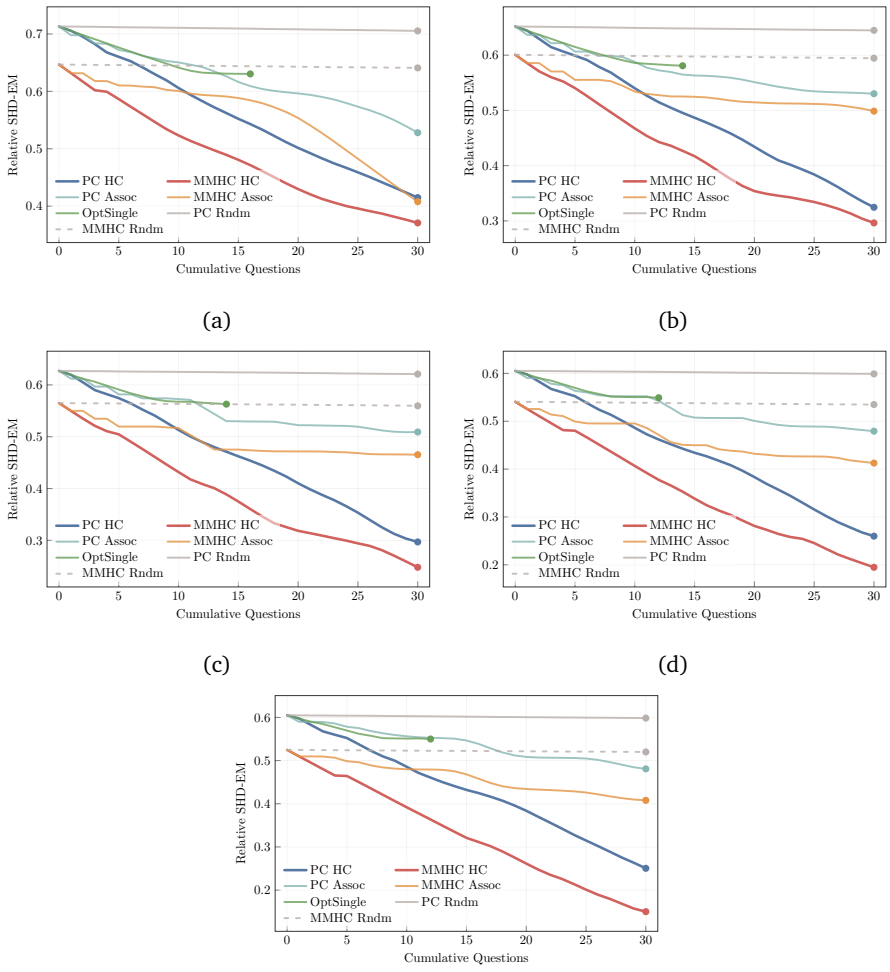

Figure 9: SHD-EM averaged over 500 replicates on the Hailfinder DAG (56 nodes, 66 edges) for sample sizes (a) $N = 500$, (b) $N = 1000$, (c) $N = 1500$, (d) $N = 3000$, (e) $N = 5000$.