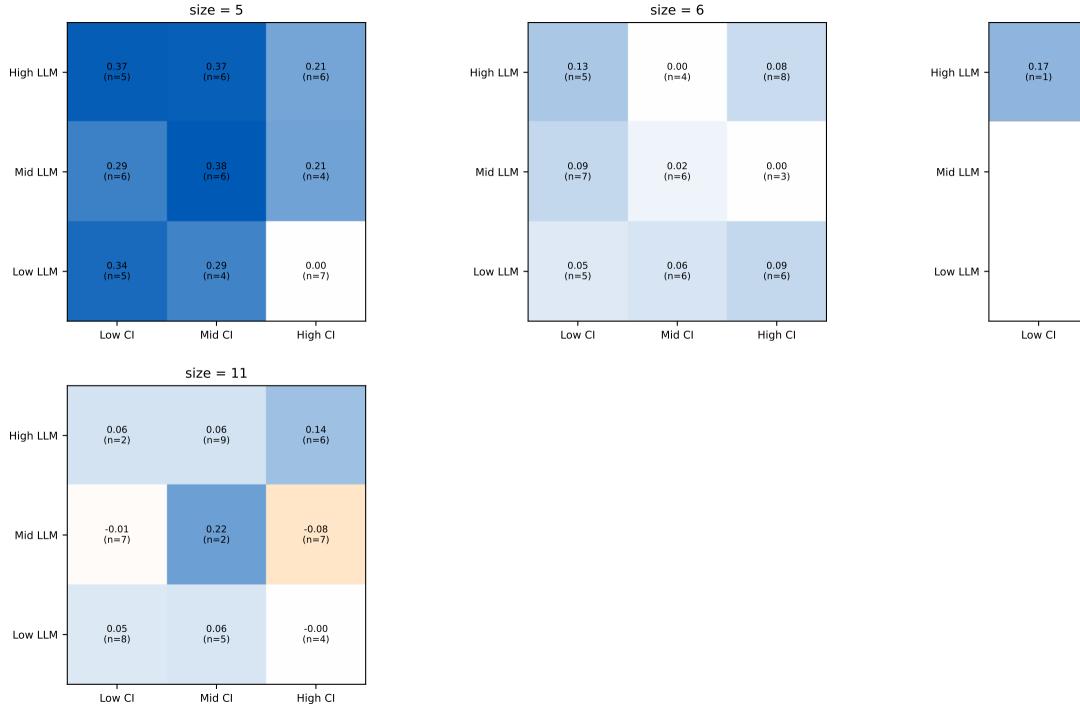
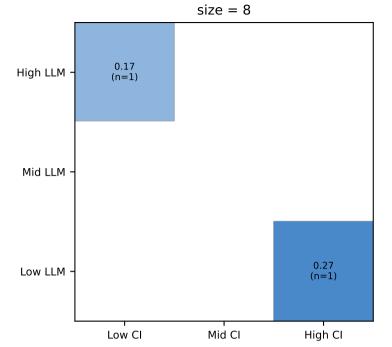


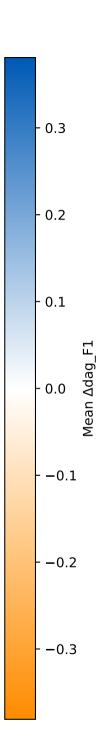
CI axis: CIT_F1 | LLM axis: prior_overall_F1

DAG Metrics

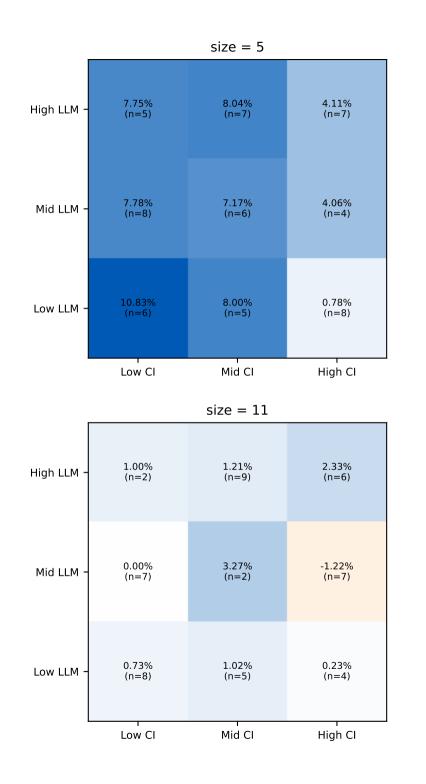
$dag_F1 - 3 \times 3$ (Low/Mid/High) by size

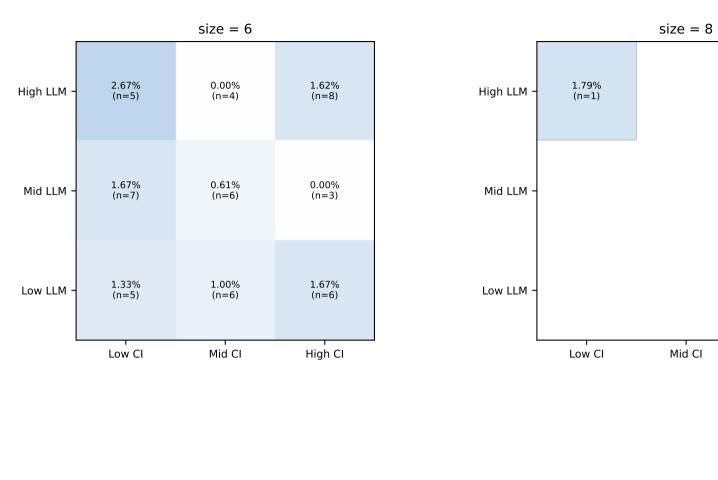


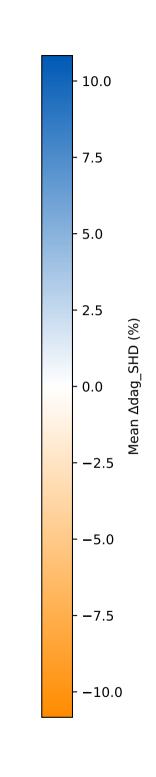




$dag_SHD - 3 \times 3$ (Low/Mid/High) by size



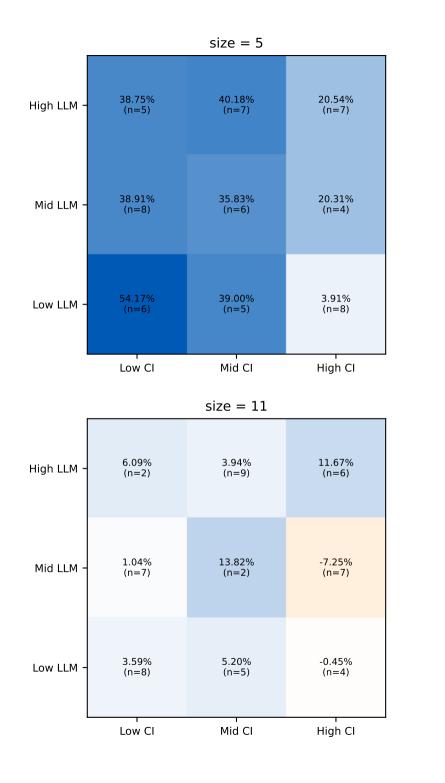


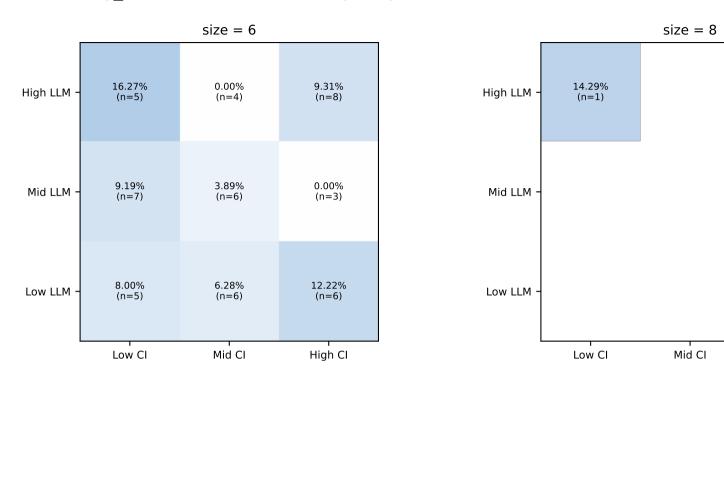


3.57% (n=1)

High CI

$dag_SID - 3 \times 3$ (Low/Mid/High) by size





- 40

- 20

Mean Adag_SID (%)

-20

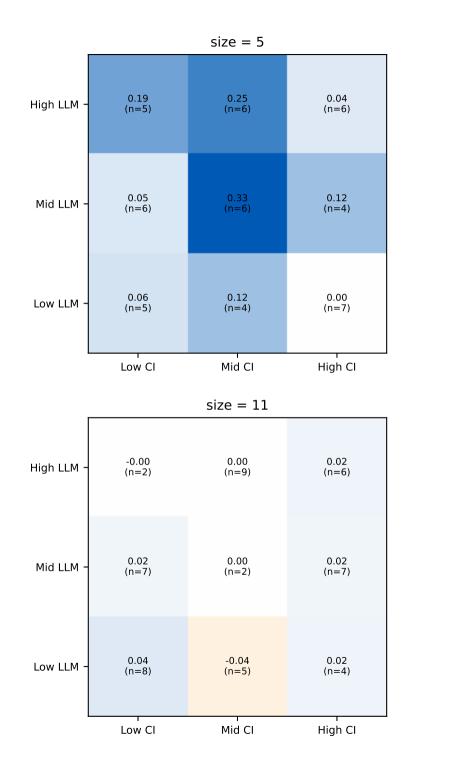
-40

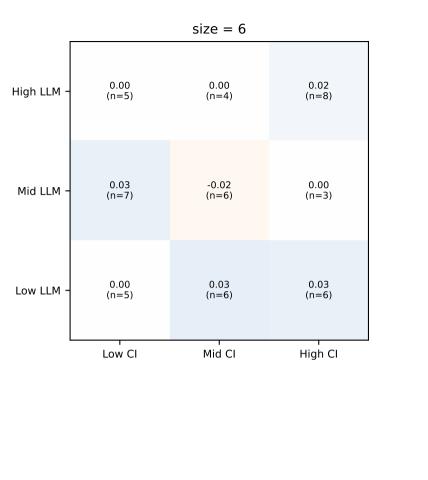
16.07% (n=1)

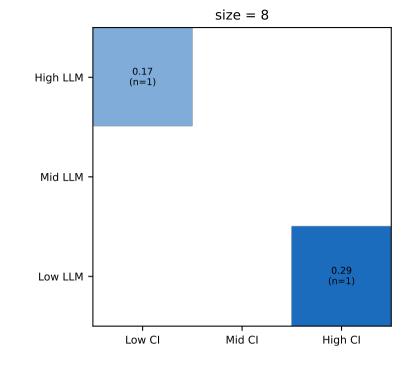
High CI

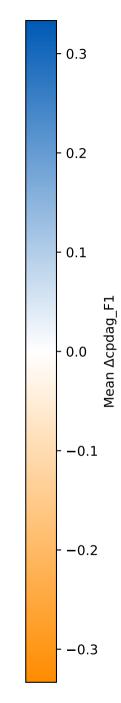
CPDAG Metrics

$cpdag_F1 - 3 \times 3$ (Low/Mid/High) by size

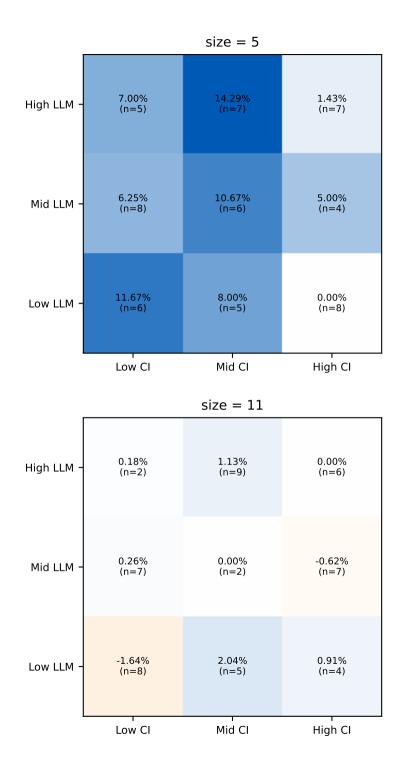


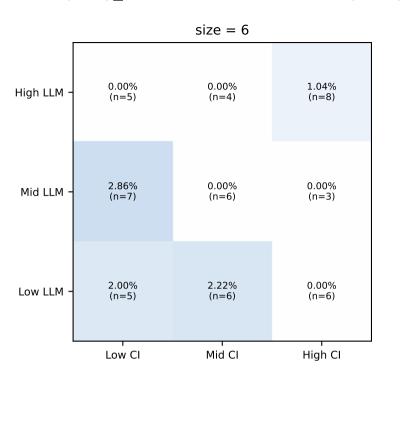


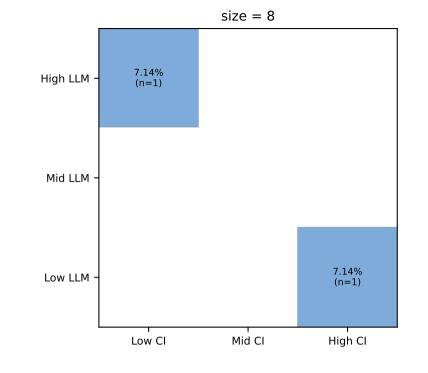


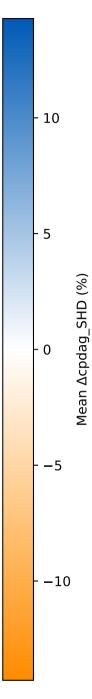


cpdag_SHD — 3×3 (Low/Mid/High) by size









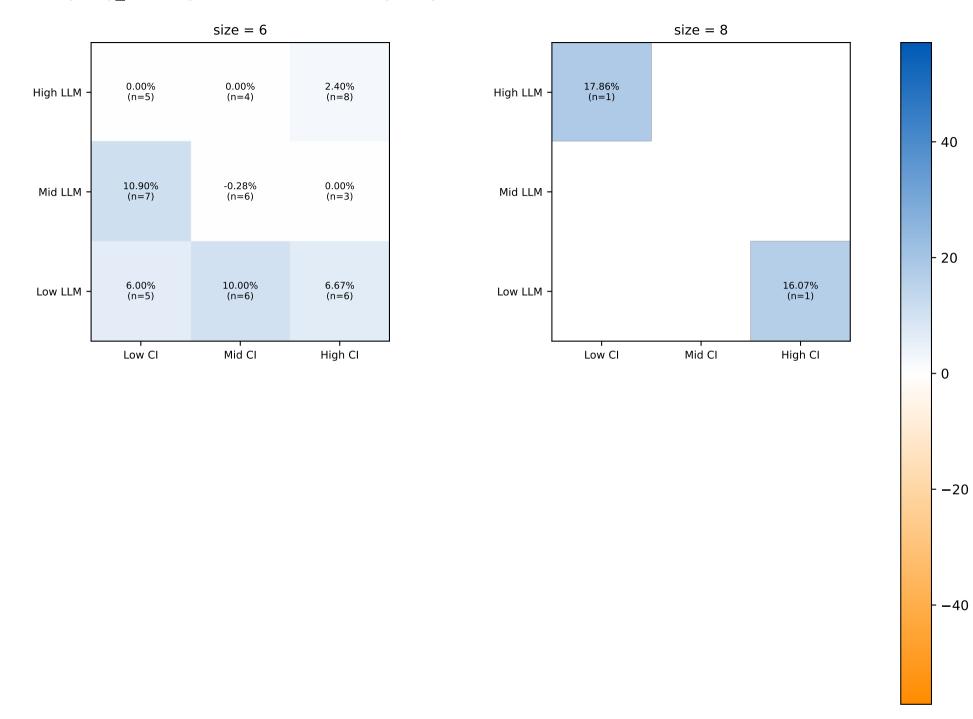
size = 510.71% (n=7) 32.50% (n=5) 57.14% (n=7) High LLM -28.75% (n=8) 48.33% (n=6) 25.00% (n=4) Mid LLM -45.83% (n=6) 40.00% (n=5) 0.00% (n=8) Low LLM -Mid CI Low CI High CI size = 11-3.36% (n=2) -0.79% (n=9) -2.70% (n=6) High LLM -1.92% (n=7) -3.64% (n=2) -3.25% (n=7) Mid LLM 3.49% (n=5) 2.55% (n=4) -3.84% (n=8) Low LLM

Low CI

Mid CI

High CI

cpdag_SID avg — 3×3 (Low/Mid/High) by size



Mean Acpdag_SID avg (%)