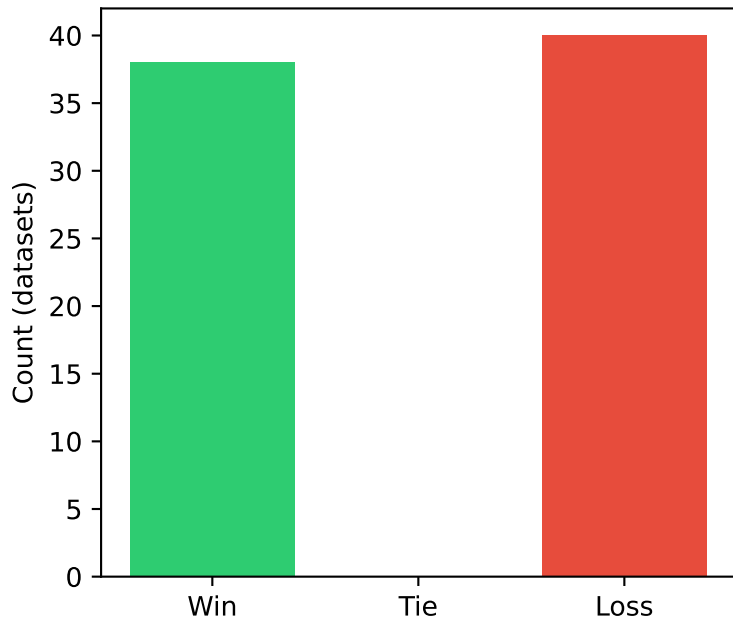


Per-dataset win/tie/loss (tie tolerance $1e-6$)

OURS vs best classical



OURS vs best GNN

