Planning Hallucination Detection: Spectral vs Baseline

- - - Baseline (LogProb) AUC=0.66
— Ours (Spectral Fiedler) AUC=0.55

False Positive Rate (Invalid labeled as Valid)

True Positive Rate (Detecting Invalid Plans)