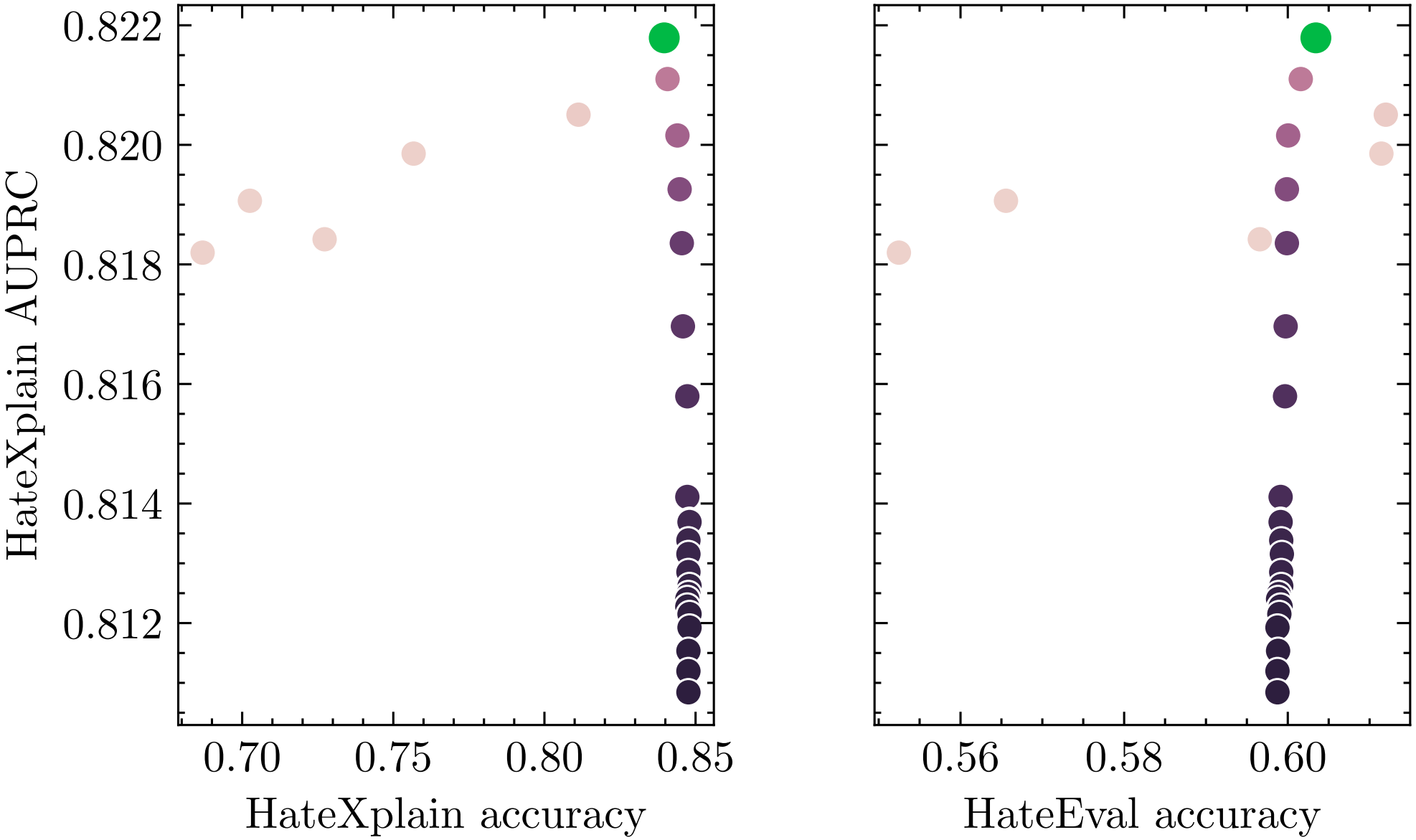
HateXplain AUPRC
0.822
0.820
0.818
0.816
0.814
0.812
0.70
0.75
0.80
0.85
HateXplain accuracy
0.56
0.58
0.60
HateEval accuracy