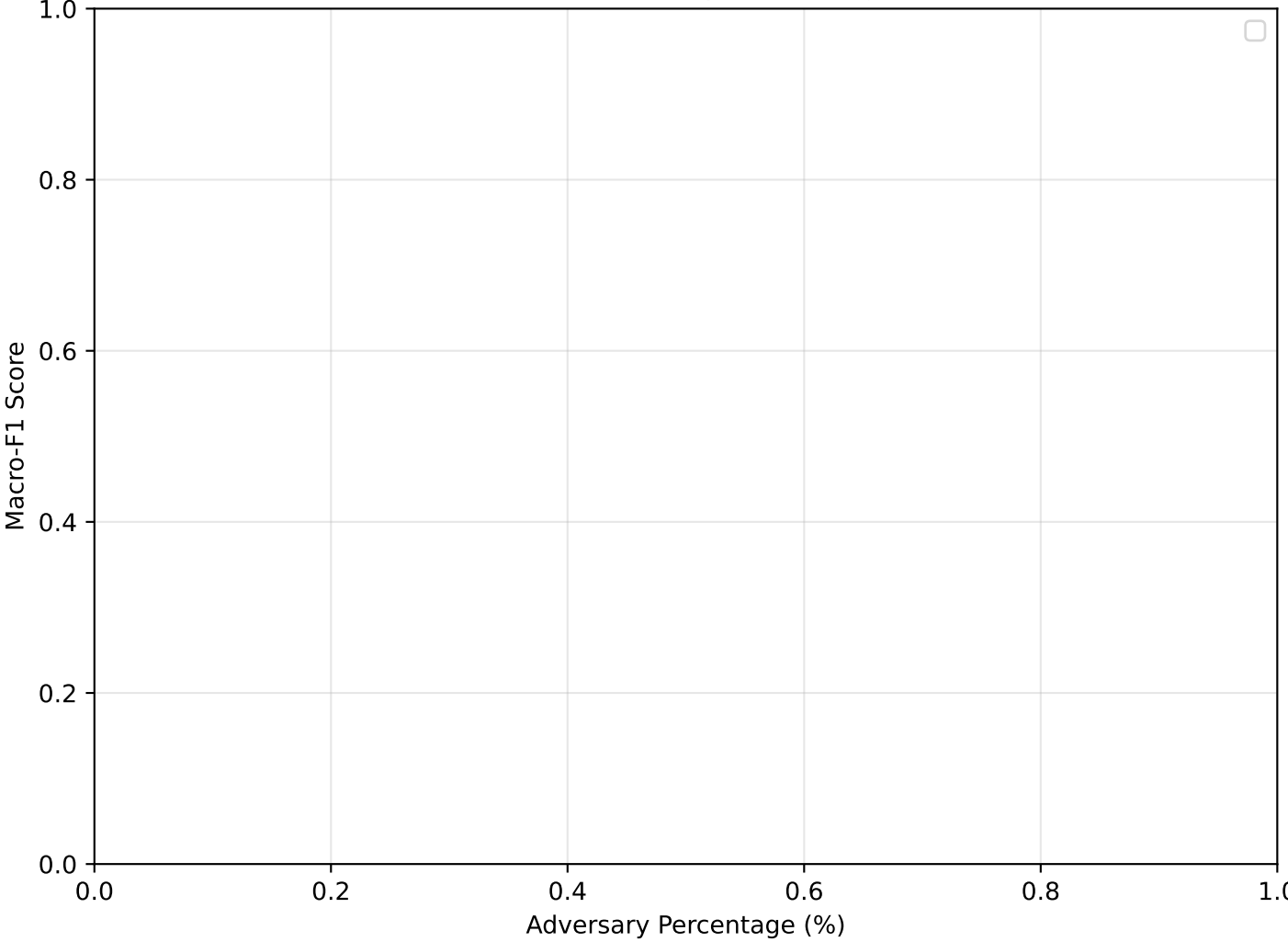
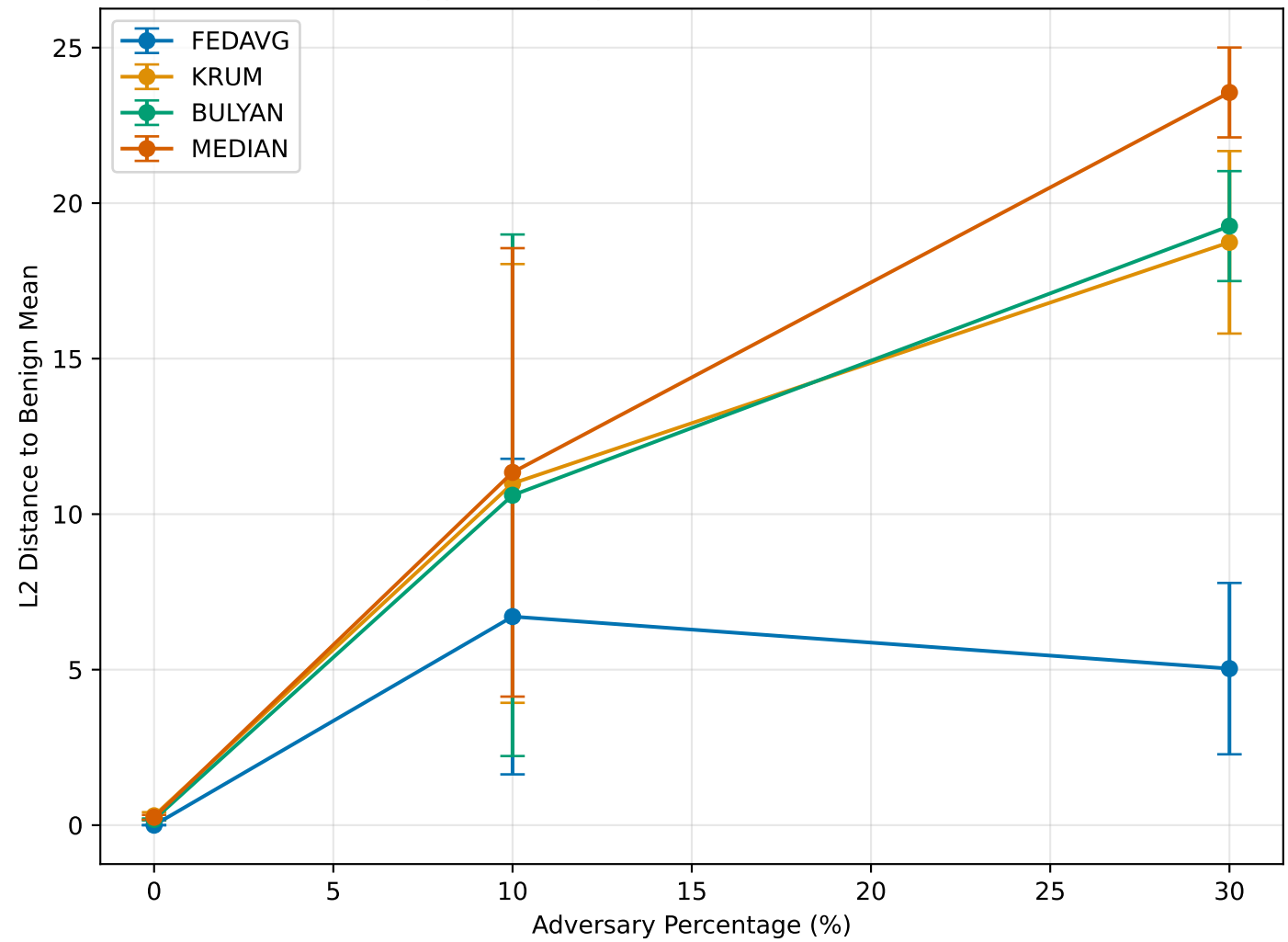


Attack Resilience Comparison
Dataset: UNSW | Clients: 6 | $\alpha=1.0$ (Dirichlet) | Attack: grad_ascent | Seeds: n=3

Detection Performance vs Adversary Percentage (Macro-F1, 95% CI)



Supplementary: Model Drift (L2 Distance)



Supplementary: Model Alignment (Cosine Similarity)

