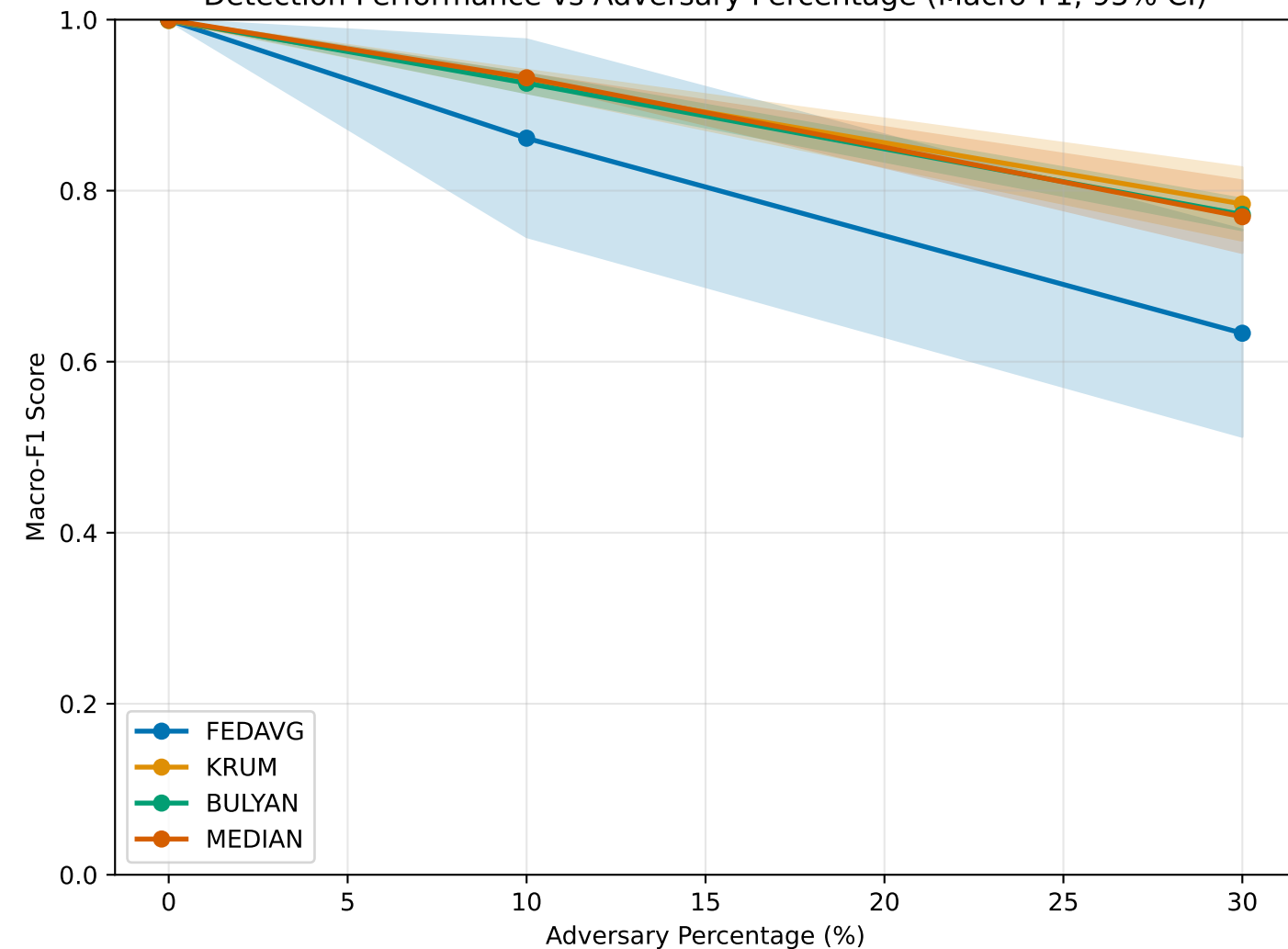


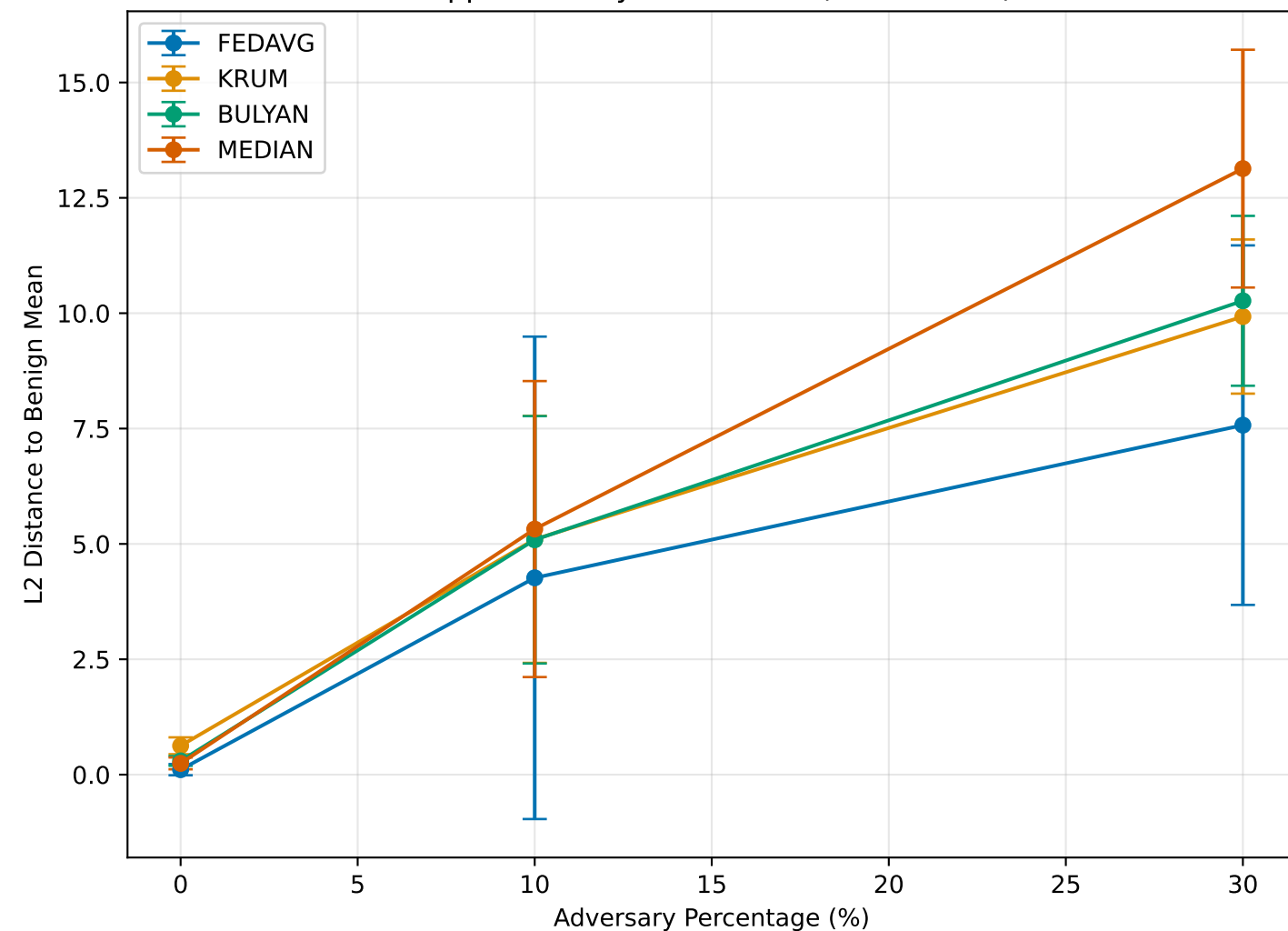
# Attack Resilience Comparison

Dataset: UNSW | Clients: 6 |  $\alpha=0.02$  (Dirichlet) | Attack: grad\_ascent | Seeds: n=5

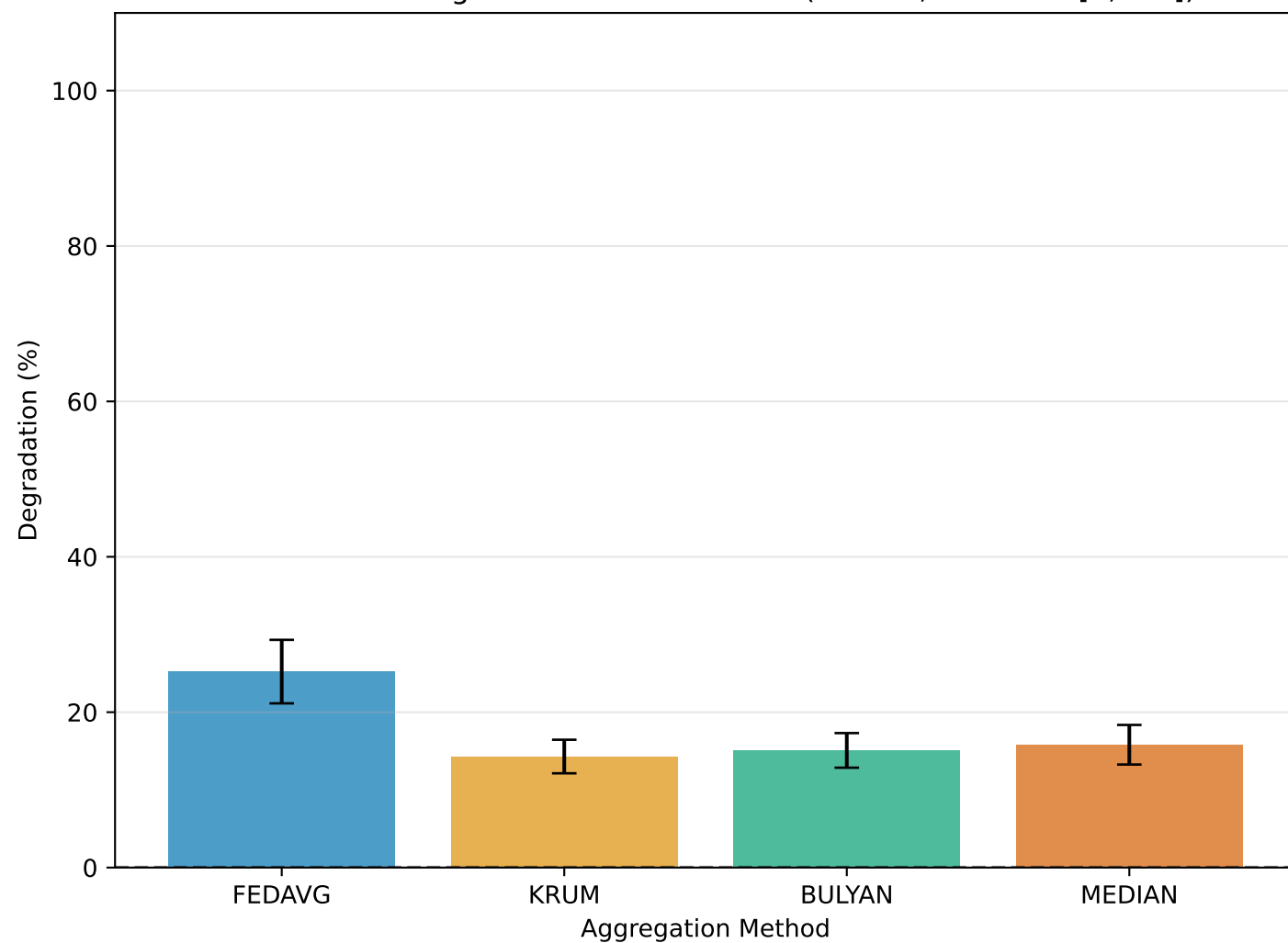
Detection Performance vs Adversary Percentage (Macro-F1, 95% CI)



Supplementary: Model Drift (L2 Distance)



Performance Degradation Under Attack (95% CI, Bounded [0,100])



Supplementary: Model Alignment (Cosine Similarity)

