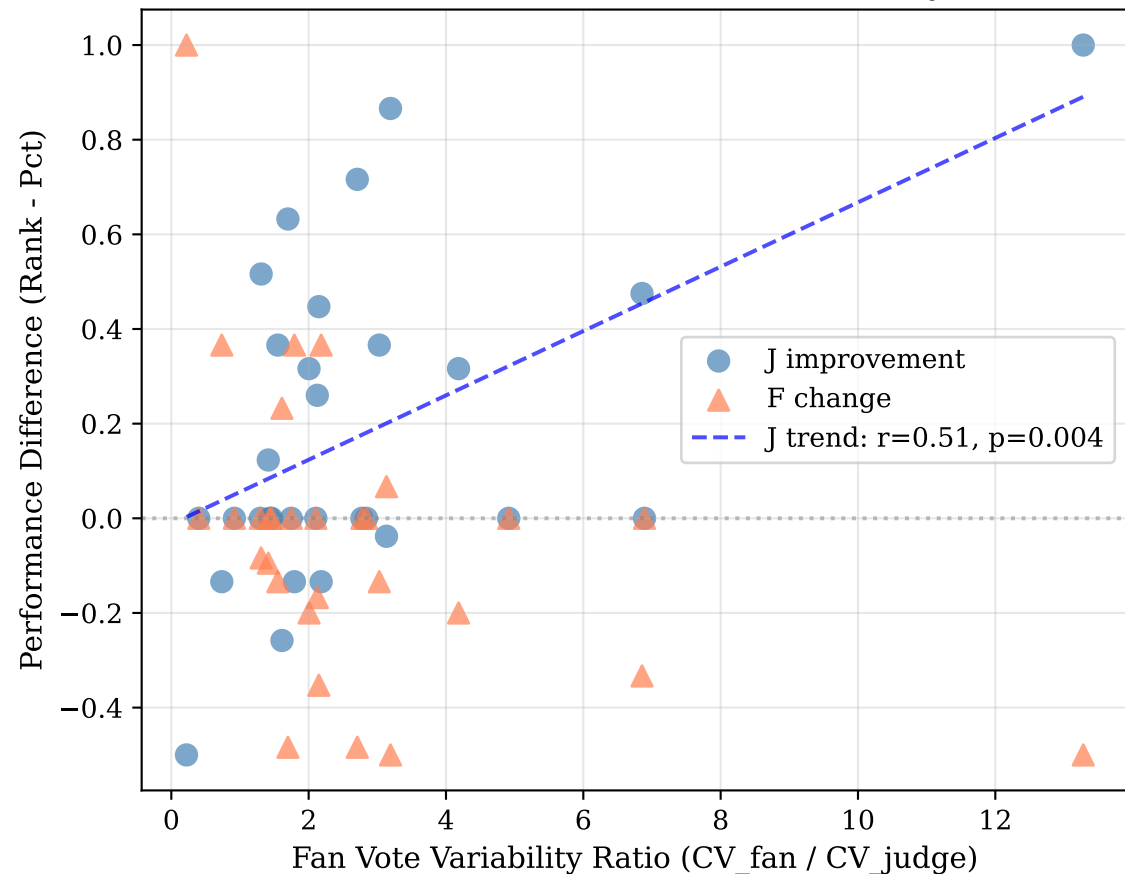(a) Robustness vs. Fan Vote Variability

(b) Performance Across Judge Systems