Stability of Measure Comparisons for scq on msmarco-subsample (p=0.5) Count Delta 0.5 600 1200 1800 2400 0.4 3000 Swap Probability 5.0 0.1 0.0 10 20 30 40 70 50 60 Average Performance Delta