Stability of Measure Comparisons for scq on msmarco-subsample (p=0.75) Count Delta 0.5 500 1000 1500 2000 0.4 2500 Swap Probability 5.0 8.0 0.1 0.0 10 20 40 30 50 60 70 Average Performance Delta