Stability of Measure Comparisons for scq on msmarco-subsample (p=0.9) Count Delta 0.5 600 1200 1800 2400 0.4 3000 Swap Probability 5.0 8.0 0.1 0.0 20 40 10 30 50 60 70 Average Performance Delta