Stability of Measure Comparisons for wig+100 on msmarco-subsample (p=0.75) Count Delta 0.5 500 1000 1500 2000 0.4 2500 Swap Probability 0.1 0.0 0.2 0.4 0.6 8.0 1.0 1.2 Average Performance Delta