Supplementary Material

Gowtham Bellala*, **Suresh K. Bhavnani**†, **Clayton Scott***
*University of Michigan, Ann Arbor, MI, †University of Texas Medical Branch, Galveston, TX
gowtham@umich.edu, skbhavnani@gmail.com, clayscot@umich.edu

Experiments

In all the figures, the plots in the first column correspond to a dataset generated using ER model, the second column correspond to a dataset generated using the PA model, the third column corresponds to the WISER database and the last column corresponds to a BRITE network.

First, we compare the performance of the rank-based algorithm with entropy-based algorithm with respect to other performance measures such as the fraction of times the unknown object θ is ranked among the top "10" in the ranked list. Here, the ranking is based on posterior probabilities where ties involving objects with equal posterior probabilities are broken randomly. Figure 1 presents these results under uniform prior for different datasets.

We then compare the performance of the rank-based algorithm with entropy-based algorithm under non-uniform prior. We consider two sets of prior: random prior where each π_i is chosen randomly from $(0, 1)$, and a Zipf's law prior where each π_i is chosen without repetition from $\{1/j\}_{j=1}^M$. The priors are then normalized such that $\sum_i \pi_i = 1$.

Figure 2 shows the results under random prior, and Figure 3 shows the results under Zipf's law prior. In both these prior models, $\frac{\rho}{1+\rho} \approx 0.04$ where $\rho = \min_i \pi_i / \max_i \pi_i$. Under the constant noise model (i.e., experiments on datasets generated using the ER model, PA model and the WISER), though the result in Proposition 1 guarantees estimation of true ranks only for $p < \frac{\rho}{1+\rho}$, we noticed that even when the noise parameter $p \gg 0.04$, the estimates of the ranks were reasonably close to the true ranks for most of the objects in Θ leading to the competitive performance of the rank-based algorithm to the entropy-based query selection as shown in Figures 2 and 3.

To compute the area under the ROC curve in these figures, the ROC curve is generated as follows: After observing responses to a set of queries, the objects are ranked based on their posterior probabilities where ties involving objects with equal posterior probabilities are broken randomly, instead of a worst case ranking. Given such a ranking of the objects in Θ, the ROC curve is obtained by varying the threshold t, where the states of the top t objects are declared as 1 and the rest 0. This leads to a certain miss rate and false alarm rate based on the rank of the unknown object θ. For example, when θ is ranked first on the list, then the ROC curve is a step function at the origin with an area equal to 1. If θ is ranked t on the list of M objects, then the ROC curve will be a step function with the transition at a false alarm rate of $(t-1)/M$.

These experiments show that though the rank-based algorithm is optimized with respect to the worst case rank, it performs competitive with the entropy-based algorithm even on other performance indicators such as "area under the ROC curve" and "the fraction of times the unknown object is among the top 10" that do not depend on the worst case rank.

Once again, it is important to note that the entropy-based algorithm is performed assuming the true knowledge of the noise parameters, whereas the rank-based algorithm does not assume this knowledge and is yet competitive with the entropy-based algorithm.

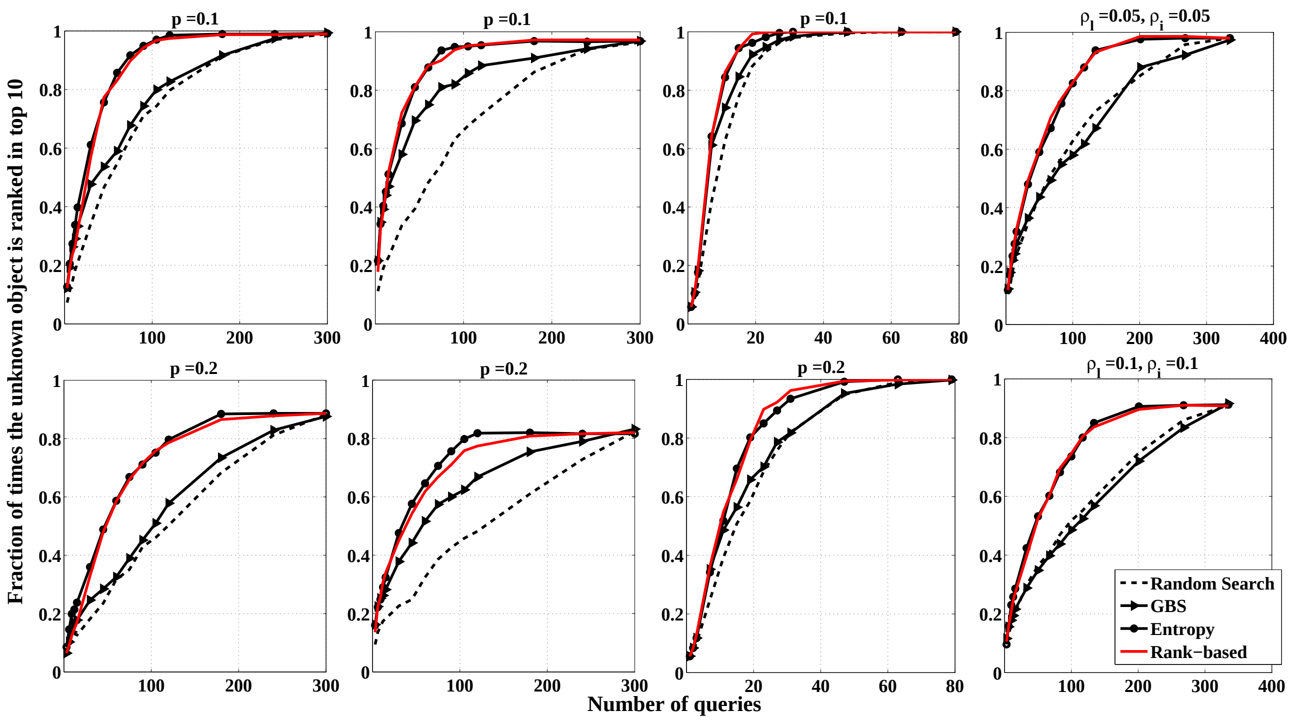

Figure 1: Demonstrates the performance of rank-based algorithm on other performance indicators that do not depend on the worst case rank. The entropy-based algorithm is implemented assuming the knowledge of the noise parameters, where as the rank-based algorithm does not use this knowledge.

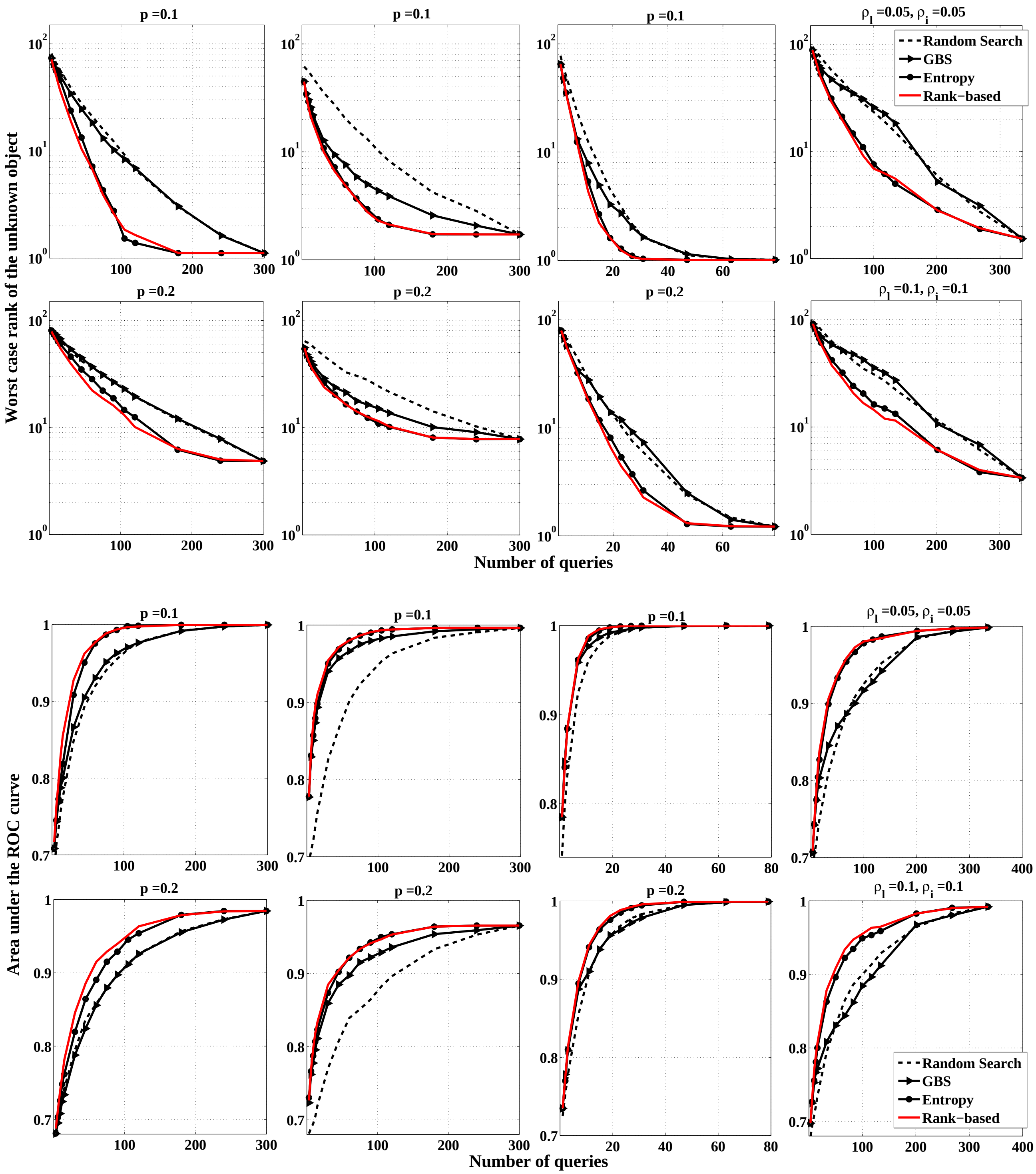

Figure 2: Performance of the rank-based algorithm under random prior. The entropy-based algorithm is implemented assuming the knowledge of the noise parameters, where as the rank-based algorithm does not use this knowledge.

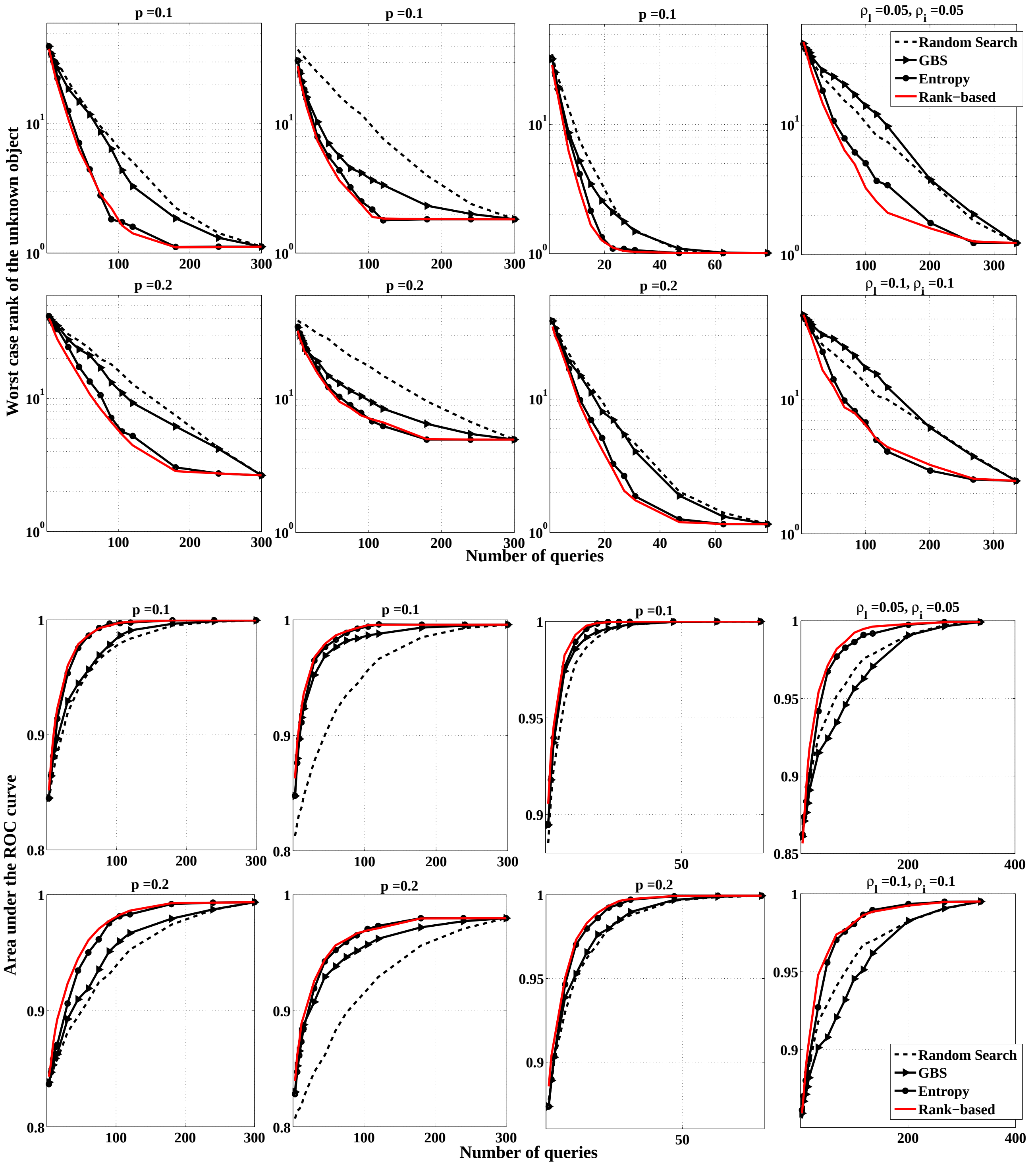

Figure 3: Performance of the rank-based algorithm under Zipf's law prior. Once again, the entropy-based algorithm is implemented assuming the knowledge of the noise parameters, where as the rank-based algorithm does not use this knowledge.