A. Missing Analysis from Section 2

Lemma 2.5. *For any nonnegative submodular function f and subset $A \subseteq N$,* UNCONSTRAINED-MAX *outputs a set $S \subseteq A$ in one adaptive round using $O(\log(1/\delta)/\varepsilon)$ oracle queries such that with probability at least $1-\delta$ we have $f(S) \ge (1/4-\varepsilon)\mathrm{OPT}_A$, where $\mathrm{OPT}_A = \max_{T\subseteq A} f(T)$.*

Proof. First assume that $\mathrm{OPT}_A > 0$. We start by bounding the individual failure probability

$$\Pr(f(R_i) \le (1/4-\varepsilon)\mathrm{OPT}_A) \le \frac{3}{3+4\varepsilon}.$$

By Lemma 2.4 we have $\mathbb{E}[f(R_i)] \ge (1/4)\mathrm{OPT}_A$. Using an analog of Markov's inequality to upper bound $\mathbb{E}[f(R_i)]$, it follows that

$$\frac{\mathrm{OPT}_A}{4} \le \mathbb{E}[f(R_i)] \le p\left(\frac{1}{4}-\varepsilon\right)\mathrm{OPT}_A + (1-p)\mathrm{OPT}_A.$$

Therefore, we must have $p \le 3/(3+4\varepsilon)$. Since the subsets R_i are chosen independently, our choice of t gives us a total failure probability of

$$\begin{aligned}\Pr(f(S) \le (1/4-\varepsilon)\mathrm{OPT}_A) &= \prod_{i=1}^{t} \Pr(f(R_i) \le (1/4-\varepsilon)\mathrm{OPT}_A)\\ &\le \left(\frac{3}{3+4\varepsilon}\right)^t\\ &\le \delta.\end{aligned}$$

This completes the proof that with probability at least $1-\delta$ we have $f(S) \ge (1/4-\varepsilon)\mathrm{OPT}_A$. To prove the adaptivity complexity, notice that all subsets R_i can be generated and evaluated at once in parallel, hence the need for only one adaptive round. For the query complexity, we use the inequality $\log(1+(4/3)\varepsilon) \ge 2\varepsilon/3$, which holds for all $\varepsilon \le 1/4$. □

B. Missing Analysis from Section 3

Corollary 3.4. *At each step of* THRESHOLD-SAMPLING *we have $\mathbb{E}[f(S_i)] \ge (1-2\hat{\varepsilon})\tau \cdot \mathbb{E}[|S_i|]$.*

Proof. We prove the claim by induction. Since f is nonnegative, the base case is clearly true. Assuming the claim as the induction hypothesis, it follows from Lemma 3.3 that

$$\begin{aligned}\mathbb{E}[f(S_{i+1})] &= \mathbb{E}[\Delta(T_{i+1}, S_i)] + \mathbb{E}[f(S_i)]\\ &\ge (1-2\hat{\varepsilon})\tau \cdot \mathbb{E}[|T_{i+1}|] + (1-2\hat{\varepsilon})\tau \cdot \mathbb{E}[|S_i|]\\ &= (1-2\hat{\varepsilon})\tau \cdot \mathbb{E}[|S_{i+1}|].\end{aligned}$$

□

Lemma 3.6. *For any subset $S \subseteq N$ and $0 \le k \le |S|$, if $T \sim \mathcal{U}(S,k)$ then $\mathbb{E}[f(T)] \ge k/|S| \cdot f(S)$.*

Proof. Fix an ordering $x_1, x_2, \dots, x_s$ on the elements in S. Expanding the expected value $\mathbb{E}[f(T)]$ and using submodularity, it follows that

$$\begin{aligned}\mathbb{E}[f(T)] &= \frac{1}{\binom{s}{k}} \sum_{R\in\mathcal{U}(S,k)} \sum_{x_i\in R} \Delta(x_i, \{x_1, x_2, \dots, x_{i-1}\} \cap R)\\ &\ge \frac{1}{\binom{s}{k}} \sum_{R\in\mathcal{U}(S,k)} \sum_{x_i\in R} \Delta(x_i, \{x_1, x_2, \dots, x_{i-1}\})\\ &= \frac{1}{\binom{s}{k}} \sum_{i=1}^{s} \binom{s-1}{k-1} \Delta(x_i, \{x_1, x_2, \dots, x_{i-1}\})\\ &= \frac{k}{s} \cdot f(S),\end{aligned}$$

which completes the proof. □

Lemma B.1. *For any set $A \subseteq N$ and optimal solution S^*, if $S_2^* = S^* \setminus A$, then*

$$f(S_2^*) - f(S_2^* \cup S) \leq f(S^*) - f(S^* \cup S).$$

Proof. It is equivalent to show that

$$f(S_2^*) + f(S^* \cup S) \leq f(S^*) + f(S_2^* \cup S).$$

For any sets $X, Y \subseteq N$, we have $f(X \cap Y) + f(X \cup Y) \leq f(X) + f(Y)$ by the definition of submodularity. It follows that

$$f(S^* \cap (S_2^* \cup S)) + f(S^* \cup S) \leq f(S^*) + f(S_2^* \cup S).$$

Therefore, it suffices to instead show that

$$f(S_2^*) \leq f(S^* \cap (S_2^* \cup S)). \tag{6}$$

Let $S_1^* = S^* \cap A$ and write

$$\begin{aligned} S^* \cap (S_2^* \cup S) &= (S^* \cap S_2^*) \cup (S^* \cap S) \\ &= S_2^* \cup (S_1^* \cap S). \end{aligned}$$

Next, fix an ordering $x_1, x_2, \dots, x_\ell$ on the elements in S^*. Summing the consecutive marginal gains of the elements in the set $S_1^* \cap S$ according to this order gives

$$f(S_2^* \cup (S_1^* \cap S)) = f(S_2^*) + \sum_{x_1, \dots, x_\ell \in S_1^* \cap S} \Delta(x_i, S_2^* \cup \{x_1, \dots, x_{i-1}\}). \tag{7}$$

We claim that each marginal contribution in (7) is nonnegative. Assume for contradiction this is not the case. Let $x^* \in S_1^* \cap S$ be the first element violating this property, and let x_{-1}^* be the previous element according to the ordering. By submodularity,

$$0 > \Delta\left(x^*, S_2^* \cup \bigcup_{x_1, \dots, x_{-1}^* \in S_1^* \cap S} \{x_i\}\right) \geq \Delta\left(x^*, S_2^* \cup \bigcup_{x_1, \dots, x_{-1}^* \in S_1^*} \{x_i\}\right),$$

which implies $f(S^* \setminus \{x^*\}) > f(S^*) = \text{OPT}$, a contradiction. Therefore, the inequality in (6) is true, as desired. □

C. Implementation Details from Section 4

We set $\varepsilon = 0.25$ for all of the algorithms except RANDOM-LAZY-GREEDY-IMPROVED, which we run with $\varepsilon = 0.01$. Since some of the algorithms require a guess of OPT, we adjust ε accordingly and fairly. We remark that all algorithms give reasonably similar results for any $\varepsilon \in [0.05, 0.50]$. We set the number of queries to be 100 for the estimators in ADAPTIVE-NONMONOTONE-MAX and BLITS, although for the theoretical guarantees these should be $\Theta(\log(n)/\varepsilon^2)$ and $\Theta(\text{OPT}^2 \log(n)/\varepsilon^2)$, respectively. For context, the experiments in (Balkanski et al., 2018) set the number of samples per estimate to be 30. Last, we set the number of outer rounds for BLITS to be 10, which also matches (Balkanski et al., 2018) since the number needed for provable guarantees is $r = 20 \log_{1+\varepsilon/2}(n)/\varepsilon$, which is too large for these datasets.