A. An Example

Suppose the ground set V consists of identical elements with equal value of 1. In other words, for any $A \subseteq V$, let $f(A)$ be 1 if A is not empty, and let it be 0 for the empty set. In this case, all elements are good candidates to be chosen at the beginning of algorithm. However after choosing any of them, the marginal gain of the rest becomes 0, and the algorithm has no incentive to continue selecting elements. If the first element is chosen deterministically, the adversary can delete that element and we can not find any non-zero value subset after deletion. Now if we pick d/ϵ of these elements and then pick one of them randomly, the probability that adversary can delete the chosen element reduces to ϵ and we achieve the robustness we aim for.

B. Explaining ROBUST-CENTRALIZED

In this section, we explain how ROBUST-CENTRALIZED returns a solution for the robust submodular maximization problem after the deletion of set D.

The subsets of A_τ and B after the deletion of set D are denoted by A'_τ and B', respectively, i.e., $A'_\tau = A_\tau \setminus D$ and $B' = B \setminus D$. ROBUST-CENTRALIZED uses the sets $\{A'_\tau\}$ and B' in order to find a good solution to the optimization problem of Eq. (1). ROBUST-CENTRALIZED considers all the possible thresholds in the range $[\Delta'_0/(2k), \Delta'_0]$, where Δ'_0 is the largest value in set $\{f(\{e\})|e \in V \setminus D\}$. We note that at this point, we can compute the value of Δ'_0 because the set of deleted elements are revealed and we also kept all elements in V_d as part of the core-set. For each threshold τ, we can ensure that the marginal gain of elements in $S_\tau = \cup_{\tau' > \tau} A_{\tau'}$ is at least τ. Therefore, we keep them as part of the solution. Next for any element $e \in B'$ the ROBUST-CENTRALIZED algorithm checks if the marginal gain of e to S_τ is at least τ. If it is, then e is added to S_τ. We do not need to introduce any extra randomness or selection from a large pool of candidates for additional robustness at this point, since the deletions are done already. The final solution is the set with the maximum value $f(S_\tau)$ among all S_τ.

C. Proof of Theorem 1

Proof. We define $V' = V \setminus D$. Assume A'_τ and B', respectively, are subsets of A_τ and B after deletion of set D from V. We define $S^* = \arg\max_{S \subseteq V \setminus D, |S| \leq k} f(S)$ and $f(S^*) = \text{OPT}$. We start by showing that one of the thresholds the ROBUST-CENTRALIZED algorithm tries is close to the standard threshold $\frac{\text{OPT}}{2k}$ that guarantees the $\frac{1}{2}$ approximation without deletion.

Lemma 1. *There is a $\tau^* \in \text{T}'$ such that $\tau^* \leq \frac{\text{OPT}}{2k} < \tau^*(1+\epsilon)$, where T' is defined in line 2 of* ROBUST-CENTRALIZED.

Proof. From the submodularity of f we have $\Delta'_0 \leq \text{OPT} \leq k\Delta'_0$. Therefore, the smallest threshold in T' is at most $\frac{\text{OPT}}{2k}$. Setting τ^* to be the largest threshold in T' that does not exceed $\frac{\text{OPT}}{2k}$ will satisfy the claim of this lemma. □

Since ROBUST-CENTRALIZED tries different thresholds and outputs the maximum value solution among them, it suffices to lower bound the expected value of $f(S_{\tau^*})$ by $(\frac{1}{2} - \delta)\text{OPT}$. We note that S_{τ^*} consists of two parts: the elements added in the first stage (ROBUST-CORESET-CENTRALIZED) that are not deleted, i.e. $\cup_{\tau \geq \tau^*} A'_\tau$, and the set of elements added in the second stage (line 7 of ROBUST-CENTRALIZED). We start by showing that the effect of deletion on the value of the first part is negligible due to the robustness of how we insert elements in ROBUST-CORESET-CENTRALIZED. To simplify the analysis, we abuse the notation, and define $A = \cup_{\tau \geq \tau^*} A_\tau$ and $A' = \cup_{\tau \geq \tau^*} A'_\tau$.

Lemma 2. $\mathbb{E}[f(A')] \geq (1-2\epsilon)\mathbb{E}[f(A)]$*, and consequently we have* $\mathbb{E}[f(S_{\tau^*})] \geq (1-2\epsilon)\mathbb{E}[f(S_{\tau^*} \cup A)]$ *where the expectations are taken over the random coin flips of* ROBUST-CORESET-CENTRALIZED.

Proof. We represent elements of A_τ with $A_\tau = \{e_{\tau,1}, \cdots, e_{\tau,n_\tau}\}$. Similarly, we define $A'_\tau = \{e'_{\tau,1}, \cdots, e'_{\tau,n'_\tau}\}$. We also define $n_\tau = |A_\tau|$ and $n'_\tau = |A'_\tau|$. We have

$$f(A) = \sum_{\tau=\tau_{max}}^{\tau^*} \sum_{l=1}^{|A_\tau|} \Delta_f(e_{\tau,l} | \cup_{\tau' > \tau} A_{\tau'} \cup \{e_{\tau,1}, \cdots, e_{\tau,l-1}\}),$$

where τ_{max} is the highest threshold in T'. The marginal gain for all elements of A_τ is sandwiched in the narrow range $[\tau, (1+\epsilon)\tau]$. Therefore, we can bound the value of A in terms of the sizes of A_τ sets and their associated thresholds:

$$\sum_{\tau \geq \tau^*} |A_\tau|\tau \leq f(A) \leq (1+\epsilon) \sum_{\tau \geq \tau^*} |A_\tau|\tau.$$

By taking the expected value of each side of these bounds, we get:

$$\sum_{\tau \geq \tau^*} \mathbb{E}[|A_\tau|]\tau \leq \mathbb{E}[f(A)] \leq (1+\epsilon) \sum_{\tau \geq \tau^*} \mathbb{E}[|A_\tau|]\tau. \tag{3}$$

Each element of A_τ is picked randomly from a set of size $\frac{d}{\epsilon}$. This means that each of these elements are deleted with a probability at most ϵ. From the submodularity of f, we know that the marginal gain of elements of A'_τ will not decrease after deletion of any other element. Note that we have $A'_\tau \subseteq A_\tau$. Therefore, we can lower bound the expected value of remaining elements, i.e., $f(A')$, similarly:

$$\begin{aligned}
\mathbb{E}[f(A')] &= \sum_{\tau=\tau_{max}}^{\tau^*} \sum_{l=1}^{|A_\tau|} \mathbb{E}[\mathbb{I}_{e_{\tau,l} \notin D} \Delta_f(e_{\tau,l} | \cup_{\tau'>\tau} A'_{\tau'} \cup \{\mathbb{I}_{e_{\tau,1} \notin D} e_{\tau,1}, \cdots, \mathbb{I}_{e_{\tau,l-1} \notin D} e_{\tau,l-1}\})] \\
&\overset{(a)}{\geq} \sum_{\tau=\tau_{max}}^{\tau^*} \sum_{l=1}^{|A_\tau|} \mathbb{E}[\mathbb{I}_{e_{\tau,l} \notin D} \Delta_f(e_{\tau,l} | \cup_{\tau'>\tau} A_{\tau'} \cup \{e_{\tau,1}, \cdots, e_{\tau,l-1}\})] \\
&\geq \sum_{\tau=\tau_{max}}^{\tau^*} \sum_{l=1}^{|A_\tau|} \Pr[e_{\tau,l} \notin D]\tau \geq (1-\epsilon) \sum_{\tau \geq \tau^*} \mathbb{E}[|A_\tau|]\tau,
\end{aligned} \tag{4}$$

where $\mathbb{I}_{e \notin D}$ is a binary indicator variable to check $e \notin D$. Inequality (a) is concluded from the submodularity of f. By combining Eqs. (3) and (4), we conclude that:

$$\mathbb{E}[f(A'] \geq \frac{1-\epsilon}{1+\epsilon} \mathbb{E}[f(A)] \geq (1-2\epsilon)\mathbb{E}[f(A)].$$

So far we have proved that the expected value of A' is not much smaller than the value of A. We note that by definition A' is a subset of both S_{τ^*} and A. By submodularity, we have:

$$f(S_{\tau^*} \cup A) - f(S_{\tau^*}) \leq f(A) - f(A').$$

We have shown that the expected value of the right hand side is at most $2\epsilon\mathbb{E}[f(A)]$ which completes the proof, since $f(A) \leq f(S_{\tau^*} \cup A)$ by monotonicity of f. □

We have shown that values of S_{τ^*} and $S_{\tau^*} \cup A$ do not differ by much. So we can focus on lower bounding $f(S_{\tau^*} \cup A)$ in the rest of the proof.

Lemma 3. $f(S_{\tau^*} \cup A) \geq \frac{(1-\epsilon)\mathrm{OPT}}{2}$.

Proof. The while loop condition in line 7 of ROBUST-CORESET-CENTRALIZED ensures that there will be at most k elements in A. If A has exactly k elements, its value is at least $k\tau^* \geq \frac{\mathrm{OPT}}{2(1+\epsilon)} \geq \frac{(1-\epsilon)\mathrm{OPT}}{2}$, since each element added to A increases its value by some threshold $\tau \geq \tau^*$. Monotonicity of f implies that $f(S_{\tau^*} \cup A) \geq f(A)$ which completes the proof in this case. Similarly, the claim is proved if S_{τ^*} has k elements. So in the rest of the proof, we focus on the case $|A| < k$ and $|S_{\tau^*}| < k$.

We define $S_{\tau^*,e}$ to be the subset of S_{τ^*} which is selected by ROBUST-CENTRALIZED exactly before processing e. We have

$$f(S^*) \overset{(a)}{\leq} f(S^* \cup S_{\tau^*} \cup A) \overset{(b)}{\leq} f(S_{\tau^*} \cup A) + \sum_{e \in S^* \setminus (S_{\tau^*} \cup A)} f(e|S_{\tau^*} \cup A)$$

$$\overset{(c)}{\leq} f(S_{\tau^*} \cup A) + \sum_{e \in (S^* \setminus (S_{\tau^*} \cup A)) \setminus B'} f(e|A) + \sum_{e \in (S^* \setminus (S_{\tau^*} \cup A)) \cap B'} f(e|S_{\tau^*,e})$$

$$\overset{(d)}{\leq} f(S_{\tau^*} \cup A) + k\tau^* \implies \frac{\text{OPT}}{2} \overset{(e)}{\leq} f(S_{\tau^*} \cup A).$$

Inequality (a) is true because f is monotone. From the submodularity of f we conclude (b). We have $A \subseteq S_{\tau^*} \cup A$ and $S_{\tau^*,e} \subseteq S_{\tau^*} \cup A$. Thus (c) results from the submodularity of f.

To prove inequality (d), we first note that the elements $e \in (S^* \setminus (S_{\tau^*} \cup A)) \setminus B'$ are discarded by ROBUST-CORESET-CENTRALIZED. Since A has strictly less than k elements, they were not discarded because of the cardinality constraint. So, for all of them we have $\Delta_f(e|A) < \tau^*$ (low marginal value). Elements $e \in (S^* \setminus (S_{\tau^*} \cup A))$ are not selected by ROBUST-CENTRALIZED, and cardinality constraint was not the reason for their rejection. Therefore, for these elements we have $f(e|S_{\tau^*,e}) < \tau^*$. □

From the results of Lemmas 2 and 3, we know $\mathbb{E}[f(S_{\tau^*})]$ is at least $\frac{(1-3\epsilon)\text{OPT}}{2}$ which proves the first claim of this theorem.

The number of thresholds in ROBUST-CORESET-CENTRALIZED is $O(\log k/\epsilon)$. For each threshold τ, we store at most d/ϵ items in a B_τ set. Also, the maximum number of elements in $\{\cup A_\tau\}$ is k. In addition, we have $d+1$ items in V_d. Therefore, the size of core-set returned by ROBUST-CORESET-CENTRALIZED is at most $O\left(k + (d \log k)/\epsilon^2\right)$ elements. For the query complexity of ROBUST-CORESET-CENTRALIZED we have: (i) each element is considered for at most $O(\log k/\epsilon)$ different thresholds, resulting in $O\left((|V| \log k)/\epsilon\right)$ oracle evaluations, and (ii) when an element is picked from B_τ to be added to A_τ, we should re-calculate marginal gain of elements and update B_τ resulting in $k|V|$ oracle evaluations since the size of the union set $\cup_{\tau \in \text{T}} A_\tau$ never exceeds k. ROBUST-CENTRALIZED receives the core-set as the input so it only processes $O\left(k + (d \log k)/\epsilon^2\right)$ elements. Each of them is considered to be added to one of the $O(\log k/\epsilon)$ sets $\{S_\tau\}_{\tau \in \text{T}'}$ which results in $O\left((k + d\frac{\log k}{\epsilon^2})\frac{\log k}{\epsilon}\right)$ oracle evaluations. □

D. The ROBUST-STREAMING Algorithm

At the end of ROBUST-CORESET-STREAMING, we know there is one running instance of the algorithm with a threshold τ^* such that $\tau^* \leq \frac{\text{OPT}}{2k} < (1+\epsilon)\tau^*$. For all $e \in V \setminus (A_{\tau^*} \cup B_{\tau^*})$, we have $\Delta_f(e|A_{\tau^*}) < \tau^*$. This ensures that the marginal gain of elements that are not picked by this running instance are smaller than $\frac{\text{OPT}}{2k}$. Let $\{A'_\tau\}$ and $\{B'_\tau\}$ be the subsets of $\{A_\tau\}$ and $\{B_\tau\}$ after the deletion of the set D from V, respectively. The elements of A_{τ^*} are robust to the deletion, i.e., $\mathbb{E}[f(A'_{\tau^*})] \geq (1-2\epsilon)\mathbb{E}[f(A_{\tau^*})]$. Also, all the elements with marginal gain of at least τ^* are kept in the set B'_{τ^*}. Finally, ROBUST-STREAMING, by adding elements of B'_{τ^*} with a marginal gain at least τ^* to A'_{τ^*}, finds a solution with an expected approximation guarantee of $\frac{1-3\epsilon}{2}$ to the optimum solution. The pseudo code of ROBUST-STREAMING is given in Algorithm 5.

E. Proof of Theorem 2

Proof. The proof is similar to the proof of Theorem 1. We define $V' = V \setminus D$. Assume A'_τ and B'_τ, respectively, are subsets of A_τ and B_τ after deletion of set D from V. We define

$$S^* = \underset{S \subseteq V \setminus D, |S| \leq k}{\arg\max} f(S) \text{ and } f(S^*) = \text{OPT}.$$

In our proof, we should consider three points. First, there is a $\tau^* \in \text{T}'$ such that $\tau^* \leq \frac{\text{OPT}}{2k} < \tau^*(1+\epsilon)$. Second, we can show that $\mathbb{E}[f(A'_{\tau^*})] \geq (1-2\epsilon)\mathbb{E}[f(A_{\tau^*})]$. Third, all the elements with enough marginal gain are in the set B'_{τ^*} and ROBUST-CENTRALIZED will add them to the final solution.

First Note that $\Delta'_0 \leq \text{OPT} \leq k\Delta'_0$ and T' contains all the thresholds in $[\frac{\Delta'_0}{2(1+\epsilon)k}, \Delta'_0]$. Also, $\Delta_d \leq \Delta'_0 \leq \Delta_0$. Therefore, there is a threshold τ^* such that $\tau^* \leq \frac{\text{OPT}}{2k} < \tau^*(1+\epsilon)$ and it is in both T' and T_n.

Algorithm 5 ROBUST-STREAMING

Input: $\{A'_\tau\}$ and $\{B'_\tau\}$ $\{A'_\tau$ and B'_τ contain elements of A_τ and B_τ (outputs of ROBUST-CORESET-STREAMING) after deletion.$\}$
Output Set S of cardinality at most k
$\Delta'_0 \leftarrow$ the largest value of set $\{f(\{e\})|e \in \{\cup A'_\tau\} \cup \{\cup B'_\tau\}\}$
$\mathrm{T}' = \{(1+\epsilon)^i | \frac{\Delta'_0}{2(1+\epsilon)k} \leq (1+\epsilon)^i \leq \Delta'_0\}$
for $\tau \in \mathrm{T}'$ **do**
  $S_\tau \leftarrow A'_\tau$
  for all $e \in B'_\tau$ **do**
    if $|S_\tau| < k$ and $\Delta_f(e|S_\tau) \geq \tau$ **then**
      $S_\tau \leftarrow S_\tau \cup e$
Return $\arg\max_\tau f(S_\tau)$

Second For the threshold τ^*, ROBUST-CORESET-STREAMING returns two sets A_{τ^*} and B_{τ^*}, where B_{τ^*} is the union of sets $B_{\tau^*,\tau}$. Assume A_{τ^*} has n_{τ^*} elements and out of these n_{τ^*} elements, $n_{\tau^*,\tau}$ elements are picked from $B_{\tau^*,\tau}$. This means their marginal gain is in the range of $[\tau, \tau(1+\epsilon)]$. We can bound $f(A_{\tau^*})$ from above by

$$\sum_{\tau \geq \tau^*} n_{\tau^*\tau}\tau \leq f(A_{\tau^*}) \leq (1+\epsilon) \sum_{\tau \geq \tau^*} n_{\tau^*\tau}\tau$$

By taking the expected value of each side of these bounds, we get:

$$\sum_{\tau \geq \tau^*} \mathbb{E}[n_{\tau^*\tau}]\tau \leq \mathbb{E}[f(A_{\tau^*})] \leq (1+\epsilon) \sum_{\tau \geq \tau^*} \mathbb{E}[n_{\tau^*\tau}]\tau \tag{5}$$

We know that an element which is picked at a given step is deleted with a probability at most ϵ. The expected number of elements picked from $B_{\tau^*,\tau}$ that remains in the set A'_τ (set A_τ after deletion) is $\mathbb{E}[n'_{\tau^*,\tau}] \geq (1-\epsilon)\mathbb{E}[n_{\tau^*,\tau}]$. Due to the submodularity of f, the marginal gain of these undeleted elements is at least τ. To sum up, we have

$$\mathbb{E}[f(A'_\tau)] \geq (1-\epsilon) \sum_{\tau \geq \tau^*} \mathbb{E}[n_{\tau^*,\tau}]\tau.$$

Therefore, we have

$$\mathbb{E}[f(A'_{\tau^*})] \geq \frac{1-\epsilon}{1+\epsilon}\mathbb{E}[f(A_{\tau^*})] \geq (1-2\epsilon)\mathbb{E}[f(A_{\tau^*})] \tag{6}$$

Let S_{τ^*} denote the set returned by ROBUST-STREAMING for threshold τ^*. To prove $\mathbb{E}[f(S_{\tau^*})] \geq (\frac{1}{2} - 3\epsilon)\mathrm{OPT}$, we consider three cases. If $|A_{\tau^*}| = k$, then $\mathbb{E}[f(A_{\tau^*})] \geq k\tau^* \geq \frac{\mathrm{OPT}}{2(1+\epsilon)} \geq (1-\epsilon)\frac{\mathrm{OPT}}{2}$ and from Eq. (6) we have $\mathbb{E}[f(S_{\tau^*})] \geq \mathbb{E}[f(A'_{\tau^*})] \geq (1-2\epsilon)\mathbb{E}[f(A_{\tau^*}) \geq \frac{(1-3\epsilon)\mathrm{OPT}}{2}$. The claim is proved similarly if S_{τ^*} has k elements. Let's assume $|A_{\tau^*}| < k$ and $|S_{\tau^*}| < k$.

Lemma 4. $\mathbb{E}[f(S_{\tau^*})] \geq (1-2\epsilon)\mathbb{E}[f(S_{\tau^*} \cup A_{\tau^*})]$. *Also if* $|A_{\tau^*}| < k$ *and* $|S_{\tau^*}| < k$, *then* $f(S_{\tau^*} \cup A_{\tau^*}) \geq \frac{(1-\epsilon)\mathrm{OPT}}{2}$.

The proof of this lemma is similar to the proofs of Lemmas 2 and 3 and we skip the details. To sum-up, for the case $|A_{\tau^*}| < k$, from Lemma 4, we have $\frac{(1-3\epsilon)\mathrm{OPT}}{2} \leq \mathbb{E}[f(S_{\tau^*})]$. This concludes the first claim of theorem.

Number of thresholds in ROBUST-CORESET-STREAMING in the interval $[\frac{\Delta_d}{2(1+\epsilon)k}, \Delta_d]$ is $O(\frac{\log k}{\epsilon})$. For each τ in this interval, there are $O(\frac{\log k}{\epsilon})$ sets of $B_{\tau,\tau'}$. We store at most $\frac{d}{\epsilon}$ elements in each of $B_{\tau,\tau'}$ set. Also, the maximum number of elements in A_τ is k. Also, there at most d elements with the marginal gain in range $(\Delta_d, \Delta_0]$. To sum up, ROBUST-CORESET-STREAMING stores $O(\frac{\log k}{\epsilon}(k + \frac{d\log k}{\epsilon^2}) + d) = O(\frac{k\log k}{\epsilon} + \frac{d\log^2 k}{\epsilon^3}))$ elements. For the time complexity of ROBUST-CORESET-STREAMING we have: (i) each element is considered in at most $O(\frac{\log k}{\epsilon})$ different thresholds resulting in $O(\frac{\log k}{\epsilon}|V|)$ oracle evaluations, and (ii) for each threshold, when an element is picked from $B_{\tau,\tau'}$ to be added to A_τ, we should re-calculate marginal gains of all elements in $\cup_{\tau'' \geq \tau} B_{\tau,\tau''}$ resulting in $O(\frac{dk\log k}{\epsilon^2})$ oracle evaluations. This is true because, for each τ, the size of A_τ never exceeds k and we have at most $O(\frac{d\log k}{\epsilon^2})$ elements in $\cup_{\tau'' \geq \tau} B_{\tau,\tau''}$. Therefore, the

total time complexity of ROBUST-CORESET-CENTRALIZED is $O(\frac{\log k}{\epsilon}|V| + \frac{dk\log^2 k}{\epsilon^3})$. ROBUST-STREAMING receives the core-set as the input so it only processes $O(\frac{k\log k}{\epsilon} + \frac{d\log^2 k}{\epsilon^3})$ elements. From the input, only $O(\frac{d\log^2 k}{\epsilon^3})$ elements are in B'_τ. Each of them is considered to be added to one of the $O(\frac{\log k}{\epsilon})$ sets $\{S_\tau\}_{\tau\in\mathrm{T}'}$ which results in $O(\frac{d\log^3 k}{\epsilon^4})$ oracle evaluations. □

F. Proof of Theorem 3

Proof. In the first round of our algorithm, we randomly distribute the elements of V on m machines. i.e., independently assigning each element to one of the m machines uniformly at random. The data assigned to machine i is represented by V_i. We also define $V' = V \setminus D$ and $V'_i = V_i \setminus D$. Let $\mathcal{V}'(1/m)$ represent the distribution over random subsets of V' where each element is sampled independently with a probability $1/m$.

Lemma 5. *The distribution of $V'_i = V_i \setminus D$ is identical to $\mathcal{V}'(1/m)$.*

Proof. Note that we assume the adversary does not have access to the randomness of our algorithm. Therefore, all the elements of $V \setminus D$ are distributed uniformly at random on m machines. □

For the sake of analysis, we assume, in each run of the algorithm, for picking elements from the pool of B_τ and tie-breaking we have a fixed strict total ordering Π of the elements of V. The choice of permutation Π is uniformly at random from the symmetric group S_n. Indeed, we assume ROBUST-CORESET-CENTRALIZED in each round among all the elements with the marginal gain of $[\tau, (1+\epsilon)\tau)$ chooses the one with the highest rank in Π. Also, we make a slight change to the algorithm: when the size of all the elements with marginal gain in a range $[\tau, (1+\epsilon)\tau)$ is exactly $\frac{d}{\epsilon}$, we choose the element with the highest priority in Π and pass all these elements to the next round (as part of the core-set). In this case, at most $d/\epsilon - 1$ elements can have a marginal gain in range $[\tau, (1+\epsilon)\tau)$. So, ROBUST-CORESET-CENTRALIZED would consider the next smaller threshold, i.e., elements with marginal gain in $[\frac{\tau}{1+\epsilon}, \tau)$.

Suppose $S^* = \arg\max_{S\subseteq V', |S|\le k} f(S)$ and $f(S^*) = \mathrm{OPT}$. In addition, let $\mathrm{OPT}_i = \max_{S\subseteq V'_i, |S|\le k} f(S)$, i.e., OPT_i is the optimum value for the data on machine i. Let's define the set O_i, conditioned on the fixed set V_i and the permutation Π, as follows

$$O_i = \{e \in S^* : e \notin \textsc{Robust-Coreset-Centralized}(V_i \cup \{e\})\}.$$

Note that while the output of ROBUST-CORESET-CENTRALIZED is random in general; if we assume the set V_i and total ordering Π are fixed a priori, then the set O_i is deterministic also.

Lemma 6. *Consider a fixed strict total ordering Π between elements of V. For all $e \in O'_i \subseteq O_i$ we have*

$$e \notin \textsc{Robust-Coreset-Centralized}(V_i \cup O'_i),$$

and

$$\textsc{Robust-Coreset-Centralized}(V_i \cup O'_i) = \textsc{Robust-Coreset-Centralized}(V_i).$$

Proof. In the first step, we show that the thresholds for ROBUST-CORESET-CENTRALIZED on sets $V_i \cup \{e\}$ and $V_i \cup O'_i$ are equal to thresholds of ROBUST-CORESET-CENTRALIZED on set V_i. First note that for $e \in O_i$, we have $f(\{e\}) \le \Delta_d$. This is true because if $f(\{e\}) > \Delta_d$, then e is picked by the algorithm as an element of the core-set (as one of the top $d+1$ singleton value elements) and it contradicts with the assumption that $e \notin \textsc{Robust-Coreset-Centralized}(V_i \cup \{e\})$. As Δ_0 and Δ_d are the same for all sets $V_i, V_i \cup \{e\}$ and $V_i \cup O'_i$, their corresponding thresholds is the same also.

We prove the equality of the output core-sets of ROBUST-CORESET-CENTRALIZED on these three different sets by induction. For this reason assume, for a threshold τ, the sets of elements chosen by ROBUST-CORESET-CENTRALIZED on both V_i and $V_i \cup O'_i$ are equal so far. We show that the two instances of algorithm pick exactly the same element in the next step. Let B_τ and B'_τ denote the set of all elements with the marginal gain in the current bucket we are processing from $\textsc{Robust-Coreset-Centralized}(V_i)$ and $\textsc{Robust-Coreset-Centralized}(V_i \cup O'_i)$, respectively. We consider two main cases. If $O'_i \cap B'_\tau = \varnothing$, then the two sets B_τ and B'_τ we are processing in the runs of the algorithm are the same. If their size is strictly less than $\frac{d}{\epsilon}$, both instances output the set $B_\tau = B'_\tau$ as part of their core-set and consider the next smaller threshold. Therefore the core-sets output by the two runs of the algorithm will remain the same in this step as well, and the

induction step is proved. Otherwise, there are at least $\frac{d}{\epsilon}$ elements, and the two instances choose the same element to add to A_τ because they take the element with the highest priority in Π.

Now consider the case $O'_i \cap B'_\tau = O''_i \neq \varnothing$. We consider two sub-cases in this part. Assume $|B_\tau| < \frac{d}{\epsilon}$, then for all $e \in O''_i$ there exists at most $\frac{d}{\epsilon} - 1$ elements in $V_i \setminus O_i$ with the marginal gain in the current bucket. This contradicts the fact that $e \in O_i$ because for every such e, the set $B_\tau \cup \{e\}$ has at most $\frac{d}{\epsilon}$ elements and therefore e will be part of the core-set.

So we can focus on the sub-case $|B_\tau| \geq \frac{d}{\epsilon}$. Since for every $e \in O''_i$, element e is not part of the core-set when added to V_i, there should be some higher priority element in B_τ than any $e \in O''_i$. This highest priority element will be picked by both runs of the algorithm. Therefore the core-sets remain the same in this step of induction as well which completes the proof. □

Next, we bound the marginal gain of elements of O_i versus elements picked from pools of $\{B^i_\tau\}$ by ROBUST-CORESET-CENTRALIZED, i.e., set A^i.

Lemma 7. *Consider a fixed strict total ordering Π between elements of V. Let A^i denote the set chosen by* ROBUST-CORESET-CENTRALIZED *on machine i. For all $e \in O_i$, we have*

1. *If $|A^i| < k$ then $\Delta_f(e|A^i) \leq \dfrac{\text{OPT}_i}{2k}$.*
2. *If $|A^i| = k$ then $\Delta_f(e|A^i) \leq \dfrac{(1+\epsilon)f(A^i)}{k} \leq \dfrac{(1+\epsilon)\text{OPT}_i}{k}$.*

Proof. From Lemma 6, we know ROBUST-CORESET-CENTRALIZED on the sets V_i and $V_i \cup O_i$ outputs the same sets.

1. From the fact that $|A^i| < k$, we conclude ROBUST-CORESET-CENTRALIZED has passed over all the thresholds and has not picked e. So we conclude $\Delta_f(e|A^i) \leq \frac{\Delta_d}{2(1+\epsilon)k} \leq \frac{\text{OPT}_i}{2k}$.
2. Denote A^i by $\{e_1, \cdots, e_{|A^i|}\}$, where e_j is the j-th element added to A^i. Also, define $A_j = \{e_1, \cdots, e_j\}$, i.e., A_j is the first j picked elements of A^i. We have

$$f(A^i) = \sum_{j=1}^{|A^i|} \Delta_f(e_j|A_{j-1}),$$

where $A_0 = \varnothing$. We know

$$\Delta_f(e|A^i) \overset{(a)}{\leq} \Delta_f(e|A_{j-1}) \overset{(b)}{\leq} (1+\epsilon)\Delta_f(e_j|A_{j-1})$$

The inequality (a) is the direct consequence of submodularity of f. We prove (b) by contradiction. Assume (b) is not true. Then e should have been taken as a part of the core-set before picking e_j, and this contradicts with e being in O_i.

To sum up, we have

$$\Delta_f(e|A^i) \leq \frac{1+\epsilon}{|A^i|} \sum_{j=1}^{|A^i|} \Delta_f(e_j|A_{j-1}) \leq \frac{(1+\epsilon)f(A^i)}{k} \leq \frac{(1+\epsilon)\text{OPT}_i}{k}.$$

□

The next step is to bound $f(O_i)$ based on $f(A^i)$ and OPT_i.

Lemma 8. $f(O_i) \leq f(A^i) + (1+\epsilon)\text{OPT}_i$.

Proof. We have

$$f(O_i) \overset{(a)}{\leq} f(O_i \cup A^i) \overset{(b)}{\leq} f(A^i) + \sum_{e \in O_i} \Delta_f(e|A^i) \overset{(c)}{\leq} f(A^i) + (1+\epsilon)\text{OPT}_i.$$

Inequality (a) drives from the monotonicity of f. Inequality (b) is true because $O_i \cap A^i = \varnothing$ and f is submodular. Inequality (c) is true from the result of Lemma 7 and the fact that $|O_i| \leq k$. □

Now, we can bound the expected value of $f(O_i)$ by the expect value of $f(S^i)$, where S^i is the result of ROBUST-CENTRALIZED from the core-set of machine i. Assume set $A^{i'}$ consists of elements of A^i after deletion. We have

$$\begin{aligned}
\mathbb{E}_\Pi[f(O_i)] &\overset{(a)}{\leq} \mathbb{E}_\Pi[A^i] + (1+\epsilon)\text{OPT}_i \\
&\overset{(b)}{\leq} \frac{(1+\epsilon)\mathbb{E}_\Pi[A^{i'}]}{1-\epsilon} + (1+\epsilon)\text{OPT}_i \\
&\overset{(c)}{\leq} \frac{(1+\epsilon)\mathbb{E}_\Pi[f(S^i)]}{1-\epsilon} + (1+\epsilon)\text{OPT}_i \rightarrow \\
(\frac{1}{3} - 2\epsilon)\mathbb{E}_\Pi[f(O_i)] &\overset{(d)}{\leq} \mathbb{E}_\Pi[f(S^i))]
\end{aligned}$$

Inequalities (a) and (b) are directly from the results of Lemma 8 and Lemma 2. We know $A^i \subseteq S^i$ and inequality (c) concludes from the monotonicity of f. Theorem 1 guarantees that ROBUST-CORESET-CENTRALIZED outputs a $(\frac{1-3\epsilon}{2}, d)$-robust randomized core-set. This ensures that, for every ground set V_i, $(\frac{1-3\epsilon}{2})\text{OPT}_i \leq \mathbb{E}_\Pi[f(S^i)]$. Inequality (d) results from this fact.

We note that the only randomness properties we need in Lemma 2 and Theorem 1 are to ensure each added element to an A set has a probability of deletion of at most ϵ with linearity of expectation. With the Π based implementation of this randomness, we achieve these properties. □

In the last step, we prove the approximation guarantee of ROBUST-DISTRIBUTED. Define vector $\boldsymbol{p}$ such that for $e \in V$, we have

$$p_e = \begin{cases} \mathbb{P}_{A\sim\mathcal{V}(1/m)}[e \in \text{ROBUST-CENTRALIZED}(A \cup \{e\})] \text{ if } e \in S^*, \\ 0 \qquad \text{otherwise.} \end{cases}$$

Lemma 9. *For $\alpha = \frac{1}{3} - 2\epsilon$ and $\beta = 1 - \frac{1}{e}$, we have*

$$\begin{aligned}
\mathbb{E}[f(S^i)] &\geq \alpha\mathbb{E}[f(O_i)] \geq \alpha f^-(\mathbf{1}_{S^*} - \boldsymbol{p}) \\
\mathbb{E}[f(T)] &\geq \beta\mathbb{E}[f(S^* \cap (\{\cup_i \cup_{\tau\in\text{T}^i} A_\tau^{i\prime}\} \cup \{\cup_i B^{i'}\})] \geq \beta f^-((\boldsymbol{p})),
\end{aligned}$$

where f^- is the Lovász extension of function f.

Proof. The proof of this lemma is similar to the proof of Barbosa et al. (2015, Theorem 5). Let Z denote the set returned by ROBUST-DISTRIBUTED. From Lemma 9, we have

$$\mathbb{E}[f(Z)] \geq \mathbb{E}[f(S^i)] \geq \alpha f^-(\mathbf{1}_{S^*} - \boldsymbol{p}) \tag{7}$$

$$\mathbb{E}[f(Z)] \geq \mathbb{E}[f(T)] \geq \beta f^-(\boldsymbol{p}) \tag{8}$$

From the result of Eqs. (7) and (8) we have

$$(\beta+\alpha)\mathbb{E}[f(Z)] \geq \alpha\beta(f^-(\mathbf{1}_{S^*} - \boldsymbol{p}) + f^-((\boldsymbol{p}))) \overset{(a)}{\geq} \alpha\beta f^-(\mathbf{1}_{S^*}) = \alpha\beta f(S^*).$$

In inequality (a), we use the convexity of Lovász extension and (Barbosa et al., 2015, Lemma 1). This proves the first part of theorem. □

From Theorem 1, we know that the size of core-set for an instance of ROBUST-CORESET-CENTRALIZED is $O(k + d\frac{\log k}{\epsilon^2})$. Therefore, the size of core-set for ROBUST-DISTRIBUTED is at most m times of this value.

Proving Corollary 1: The first part of Corollary 1 is a direct consequence of Theorem 1. The second part results from the approximation guarantees of Theorems 1 and 3.

G. Experimental Results: Fairness in Crime Rate Prediction

In the second experiment for robust feature selection, we use the *Communities and Crime* dataset from UCI Repository of machine learning databases (Blake & Merz, 1998). This dataset consist of 122 features with plausible connection to crime in communities within the United States. The crime rate is provided as the per capita violent crimes. In this experiment, we delete sensitive features such as distribution of race and sex in population and police forces. Fig. 3 compares normalized objective values for $k \in \{4, 5\}$ and different number of deletions. Again, we observe that our centralized and streaming algorithms have the best performances. We should point out that the parameter d can also play an important role in practice. Indeed, since all algorithms are made robust to deletion of $d = 3$ elements, the performance of ROBUST (Mirzasoleiman et al., 2017) hugely decreases with only $r = 4$ deletions, while our algorithms maintain their near optimal performances.

To assess the quality of selected features, we use a RIDGE regression classifier (Hoerl & Kennard, 1970). From Table 3 we observe that the RMSE for a classifier that is trained on all features is 0.136. For classifiers trained on features selected by GREEDY and GREEDY$_D$, the errors increase to 0.193 and 0.199, respectively. The errors for centralized (0.163) and streaming (0.177) algorithms are even less than the greedy algorithm which knows the deleted features in advance. This might be due to the fact that only our proposed methods select features related to the percentage of divorced males and females as important attributes. It is plausible that these attributes can have high correlations with crime rate.

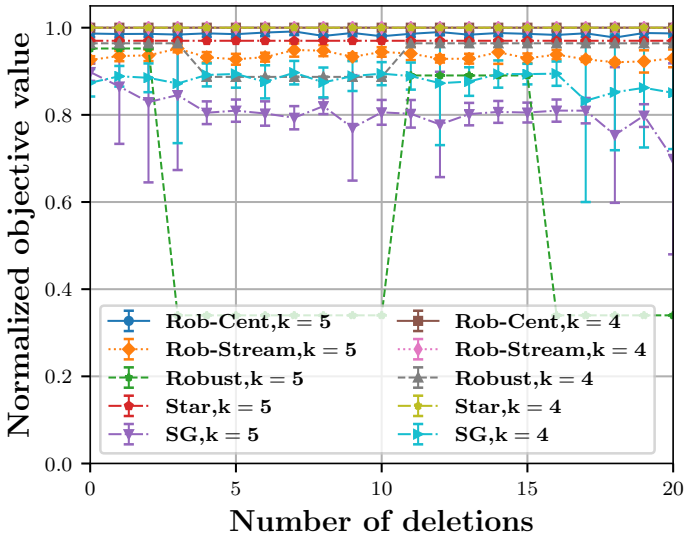

Figure 3. Crime and Communities: the effect of deletion on the performance of different algorithms for feature selection. We set $d = 3$

Table 3. The comparison of the RIDGE classifier for Crime and Communities dataset. Ten sensitive features are deleted. The number of stored features is reported in parenthesis.

Algorithm	RIDGE (RMSE)
All features	0.136
GREEDY	0.193
GREEDY$_D$	0.199
Rob-Cent	0.163 (25)
Rob-Stream	0.177 (52)
ROBUST	0.197 (58)
STAR-T-GREEDY	0.173 (71)