A. Summary of Notations

Table 2 provides the summary of notations used in this paper.

Table 2. Summary of important terminology

Set	Cardinality	Description
F	m	Set of functions $(f_1, \dots, f_m)$ drawn from an unknown distribution $\mathbb{D}$ of monotone submodular functions.
Ω	n	Given ground set of all elements. Generally so large that even greedy is too expensive.
$S^{m,\ell}$	ℓ	The optimum solution to Eq. (2), i.e., $S^{m,\ell} = \arg\max_{S\subseteq\Omega,\|S\|\leq\ell} \frac{1}{m}\sum_{i=1}^{m} \max_{\|T\|\leq k, T\subseteq S} f_i(T)$.
$S_i^{m,\ell}$	k	The optimum solution to each function f_i from set $S^{m,\ell}$, i.e., $S_i^{m,\ell} = \arg\max_{S\subseteq S^{m,\ell},\|S\|\leq k} f_i(S)$.
OPT	1	The value of optimum solution to Eq. (2), i.e., $\text{OPT} = \frac{1}{m}\sum_{i=1}^{m} f_i(S_i^{m,\ell})$.
S	ℓ	Reduced subset of elements we want to select. Ideally sublinear in n, but still representative.
T_i	k	Solution we select for each function f_i (chosen from S), i.e., $T_i \subset S$.

B. Proof of Theorem 1

Let S^t represent the set of chosen elements at step t. Also, we define $T_i^t \subseteq S^t$ as the current solution for function f_i at step t. We also define $A_i^t = \bigcup_{1\leq j\leq t} T_i^t$, i.e., A_i^t is the set of all the elements have been in the set T_i till step t. Note that this set includes elements that have been in T_i at some point and might be deleted at later steps. We first lower bound $f_i(T_i^t)$ based on value of $f_i(A_i^t)$.

Lemma 3. *For all $1 \leq i \leq m$, we have*

$$f_i(T_i^t) \geq \frac{\alpha}{\alpha+1} f_i(A_i^t).$$

Proof. We proof this lemma by induction. For the first k additions to set T_i^t, the two sets T_i^t and A_i^t are exactly the same, i.e., we have $f_i(T_i^t) = f_i(A_i^t)$. Therefore the lemma is correct for them. Next we show that lemma is correct for cases after the first k additions, i.e., when an incoming element u^t replaces one element of T_i^{t-1}. We have the following lemma.

Lemma 4. *For $1 \leq i \leq m$ and all u^t, we have:*

$$\Delta_i(u^t, T_i^{t-1}) \geq f_i(u^t|A_i^{t-1}) - f_i(T_i^{t-1})/k.$$

Proof. To prove this lemma we have the following

$$\begin{aligned}
\Delta_i(u^t, T_i^{t-1}) &= f_i(T_i^{t-1} + u^t - \textsc{Rep}_i(u^t, T_i^{t-1})) - f_i(T_i^{t-1}) \\
&\overset{(a)}{\geq} \frac{\sum_{u\in T_i^{t-1}} f_i(T_i^{t-1} + u^t - u)) - f_i(T_i^{t-1})}{k} \\
&= \frac{\sum_{u\in T_i^{t-1}} f_i(T_i^{t-1} + u^t - u) - f_i(T_i^{t-1} - u) + f_i(T_i^{t-1} - u) - f_i(T_i^{t-1})}{k} \\
&\overset{(b)}{\geq} \frac{\sum_{u\in T_i^{t-1}} f_i(T_i^{t-1} + u^t) - f_i(T_i^{t-1})}{k} + \frac{\sum_{u\in T_i^{t-1}} f_i(T_i^{t-1} - u) - f_i(T_i^{t-1})}{k} \\
&\overset{(c)}{\geq} f_i(u^t|T_i^{t-1}) - f_i(T_i^{t-1})/k \overset{(d)}{\geq} f_i(u^t|A_i^{t-1}) - f_i(T_i^{t-1})/k.
\end{aligned}$$

Inequality (a) is true because $\textsc{Rep}_i(u^t, T_i^{t-1})$ is the element with the largest increment when it is exchanged with u^t. Therefore, it should be at least equal to the average of all possible exchanges. Note that T_i^{t-1} has at most k elements. Inequalities (b) and (d) result from submodularity of f_i. Also, from submodularity of f_i, we have $f_i(T_i^{t-1}) - f_i(\varnothing) \geq \sum_{u\in T_i^{t-1}} f_i(T_i^{t-1}) - f_i(T_i^{t-1} - u)$ which results in inequality (c). $\square$

Now, assume Lemma 3 is true for time $t-1$, i.e., $f_i(T_i^{t-1}) \geq \frac{\alpha}{\alpha+1} f_i(A_i^{t-1})$. We prove that it is also true for time t. First note that if u^t is not accepted by the algorithm for the i-th function then $T_i^t = T_i^{t-1}$ and $A_i^t = A_i^{t-1}$; therefore the lemma is true for t. If u^t is chosen to be added to T_i^{t-1}, from the definition of $\nabla(u^t, T_i^{t-1})$, we have $\Delta_i(u^t, T_i^{t-1}) > \alpha/k \cdot f_i(T_i^{t-1})$. From this fact and Lemma 4, we have:

$$\begin{aligned} f_i(T_i^t) - f_i(T_i^{t-1}) &\geq \max\{f_i(u^t|A_i^{t-1}) - f_i(T_i^{t-1})/k, \alpha/k \cdot f_i(T_i^{t-1})\} \\ &\geq \frac{\alpha \cdot (f_i(u^t|A_i^{t-1}) - f_i(T_i^{t-1})/k) + \alpha/k \cdot f_i(T_i^{t-1})}{\alpha+1} \\ &\geq \frac{\alpha}{\alpha+1} \cdot f_i(u^t|A_i^{t-1}) = \frac{\alpha}{\alpha+1} \cdot \left[f_i(A_i^t) - f_i(A_i^{t-1})\right] \rightarrow f_i(T_i^t) \geq \frac{\alpha}{\alpha+1} \cdot f_i(A_i^t). \end{aligned}$$

□

Corollary 1. *If $\Delta_i(u^t, T_i^{t-1}) < \alpha/k \cdot f_i(T_i^{t-1})$ then we have:*

$$f_i(u^t|A_i^n) \overset{(a)}{\leq} f_i(u^t|A_i^{t-1}) \overset{(b)}{\leq} \frac{\alpha+1}{k} \cdot f_i(T_i^{t-1}) \overset{(c)}{\leq} \frac{\alpha+1}{k} \cdot f_i(T_i^n).$$

Proof. Inequality (a) is true because of submodularity of f_i and the fact that $A_i^{t-1} \subseteq A_i^n$. Inequality (b) concludes form Lemma 4. Since $f_i(T_i^t)$ is a nondecreasing function of t, then (c) is true. □

Next, we use Lemmas 3 and 4 and Corollary 1, to prove the approximation factor of the algorithm. Note that if at the end of algorithm $|S^n| = \ell$, then we have:

$$\frac{1}{m}\sum_{i=1}^m f_i(T_i^n) = \frac{1}{m}\sum_{t=1}^n\sum_{i=1}^m \left[f_i(T_i^t) - f_i(T_i^{t-1})\right] = \frac{1}{m}\sum_{t=1}^n \mathbb{1}_{\{u^t\in S^n\}} \cdot \nabla_i(u^t, T_i^t) \geq \frac{\text{OPT}}{\beta}. \tag{5}$$

This is true because the additive value after adding an element to S^t is at least $\frac{\text{OPT}}{\beta\ell}$. Next consider the case where $|S| < \ell$. First note that for an element $u^t \in S_i^{m,\ell}$, which does not belong to set A_i^n, we have two different possibilities: (i) $\Delta_i(u^t, T_i^{t-1}) < \alpha/k \cdot f_i(T_i^{t-1})$, or (ii) $\Delta_i(u^t, T_i^{t-1}) \geq \alpha/k \cdot f_i(T_i^{t-1})$ and $\frac{1}{m}\sum_{i=1}^m \nabla_i(u^t, T_i^{t-1}) < \frac{\text{OPT}}{\beta\ell}$. Therefore, we have

$$\begin{aligned} \sum_{i=1}^m f_i(S_i^{m,\ell}) &\leq \sum_{i=1}^m \left[f_i(A_i^n) + \sum_{u^t\in S_i^{m,\ell}\setminus A_i^n} f_i(u^t|A_i^n)\right] \\ &= \sum_{i=1}^m f_i(A_i^n) + \sum_{i=1}^m \sum_{u^t\in S^{m,\ell}\setminus A_i^n} \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}} \cdot f(u^t|A_i^n) \\ &= \sum_{i=1}^m f_i(A_i^n) + \sum_{i=1}^m \sum_{u^t\in S^{m,\ell}} \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}} \cdot \\ &\quad \left[\mathbb{1}_{\{\Delta_i(u^t,T_i^{t-1})<\alpha/k\cdot f_i(T_i^{t-1})\}} \cdot f_i(u^t|A_i^n) + \mathbb{1}_{\{\Delta_i(u^t,T_i^{t-1})\geq\alpha/k\cdot f_i(T_i^{t-1}) \text{ and } \sum_{i=1}^m \nabla_i(u^t,T_i^{t-1})<\frac{\text{OPT}}{\beta\ell}\}} \cdot f_i(u^t|A_i^n)\right]. \end{aligned} \tag{6}$$

For the three terms on the rightmost side of Eq. (6) we have the following inequalities. For the first term, from Lemma 3, we have:

$$\sum_{i=1}^m f_i(A_i^n) \leq \frac{\alpha+1}{\alpha}\sum_{i=1}^m f_i(T_i^n). \tag{7}$$

For the second term, we have:

$$\begin{aligned} &\sum_{i=1}^m \sum_{u^t\in S^{m,\ell}} \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}} \cdot \mathbb{1}_{\{\Delta_i(u^t,T_i^{t-1})<\alpha/k\cdot f_i(T_i^{t-1})\}} \cdot f_i(u^t|A_i^n) \\ &\qquad \overset{(a)}{\leq} \sum_{i=1}^m \sum_{u^t\in S_i^{m,\ell}} \frac{\alpha+1}{k} f_i(T_i^n) \overset{(b)}{\leq} (\alpha+1) \cdot \sum_{i=1}^m f_i(T_i^n). \end{aligned} \tag{8}$$

Inequality (a) is the result of Corollary 1. Inequality (b) is true because we have at most k elements in set $S_i^{m,\ell}$. Note that for u^t with $\sum_{i=1}^m \nabla_i(u^t, T_i^{t-1}) < \frac{\text{OPT}}{\beta\ell}$ we have:

$$\frac{1}{m}\sum_{i=1}^m \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}}\cdot \mathbb{1}_{\{\Delta_i(u^t,T_i^{t-1})\geq \alpha/k\cdot f_i(T_i^{t-1})\}}\left[f_i(u^t|A_i^{t-1}) - f_i(T_i^{t-1})/k\right]$$
$$\overset{(a)}{\leq} \frac{1}{m}\sum_{i=1}^m \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}}\nabla_i(u^t,T_i^{t-1}) \overset{(b)}{\leq} \frac{1}{m}\sum_{i=1}^m \nabla_i(u^t,T_i^{t-1}) < \frac{\text{OPT}}{\beta\ell}. \tag{9}$$

Inequality (a) results from Lemma 4 and (b) is true because $\nabla_i(u^t, T_i^{t-1}) \geq 0$ for $1 \leq i \leq m$. Therefore, from Eq. (9) and submodularity of f_i and its non-negativity, we have:

$$\frac{1}{m}\sum_{i=1}^m \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}}\cdot \mathbb{1}_{\{\Delta_i(u^t,T_i^{t-1})\geq \alpha/k\cdot f_i(T_i^{t-1}) \text{ and } \sum_{i=1}^m \nabla_i(u^t,T_i^{t-1})<\frac{\text{OPT}}{\beta\ell}\}}\cdot f_i(u^t|A_i^n)$$
$$\leq \frac{\text{OPT}}{\beta\ell}+\frac{1}{km}\cdot\sum_{i=1}^m \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}}\cdot f_i(T_i^n).$$

Consequently,

$$\frac{1}{m}\sum_{i=1}^m \sum_{u^t\in S^{m,\ell}} \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}}\cdot \mathbb{1}_{\{\Delta_i(u^t,T_i^{t-1})\geq \alpha/k\cdot f_i(T_i^{t-1}) \text{ and } \sum_{i=1}^m \nabla_i(u^t,T_i^{t-1})<\frac{\text{OPT}}{2\ell}\}}\cdot f_i(u^t|A_i^n)$$
$$\leq \frac{1}{m}\sum_{u^t\in S^{m,\ell}}\left[\frac{\text{OPT}}{\beta\ell}+\frac{1}{k}\cdot\sum_{i=1}^m \mathbb{1}_{\{u^t\in S_i^{m,\ell}\}}\cdot f_i(T_i^n)\right] \leq \frac{\text{OPT}}{\beta}+\frac{1}{m}\sum_{i=1}^m f_i(T_i^n). \tag{10}$$

Using Eqs. (7), (8) and (10) we have:

$$\text{OPT} = \frac{1}{m}\sum_{i=1}^m f_i(S_i^{m,\ell}) \leq \frac{\alpha+1}{\alpha}\cdot\frac{1}{m}\sum_{i=1}^m f_i(T_i^n) + (\alpha+1)\cdot\frac{1}{m}\sum_{i=1}^m f_i(T_i^n) + \frac{\text{OPT}}{\beta} + \frac{1}{m}\sum_{i=1}^m f_i(T_i^n). \tag{11}$$

This results in

$$\frac{\alpha\cdot(\beta-1)\cdot\text{OPT}}{\beta\cdot((\alpha+1)^2+\alpha)} \leq \frac{1}{m}\sum_{i=1}^m f_i(T_i^n). \tag{12}$$

Combination of Eqs. (5) and (12) proves the theorem.

C. Proof of Theorem 2

We first prove Lemmas 1 and 2.

Proof of Lemma 1: The lower bound is trivial. For the upper bound we have

$$\text{OPT} = \frac{1}{m}\sum_{i=1}^m \sum_{u\in S_i^{m,\ell}} f_i(u) \leq \frac{1}{m}\sum_{u\in S^{m,\ell}}\sum_{i=1}^m f_i(u) \leq \ell\cdot\delta.$$

Proof of Lemma 2: We have

$$\frac{1}{m}\sum_{i=1}^m \nabla_i(u^t,T_i^{t-1}) \overset{(a)}{\leq} \frac{1}{m}\sum_{i=1}^m f_i(u^t|T_i^{t-1}) \overset{(b)}{\leq} \frac{1}{m}\sum_{i=1}^m f_i(u^t) \overset{(c)}{\leq} \delta_t.$$

For inequality (a) first note that $f_i(u^t|T_i^{t-1}) \geq 0$; therefore it suffices to show that for all $\nabla_i(u^t, T_i^{t-1}) > 0$ we have $\nabla_i(u^t, T_i^{t-1}) \leq f_i(u^t|T_i^{t-1})$. So, for $\nabla_i(u^t, T_i^{t-1}) > 0$, consider the two following cases: (i) if $|T_i^{t-1}| < k$, then

$\nabla_i(u^t, T_i^{t-1}) = f_i(u^t|T_i^{t-1})$. (ii) if $|T_i^{t-1}| < k$, then $\nabla_i(u^t, T_i^{t-1}) = \Delta_i(u^t, T_i^{t-1}) = f_i(T_i^{t-1} + u^t - \textsc{Rep}_i(u^t, T_i^{t-1})) - f_i(T_i^{t-1}) \leq f_(T_i^{t-1} + u^t) - f_i(T_i^{t-1})$, where the last inequality follows from the monotonicity of f_i. Inequality (b) results from the submodularity of f_i. The inequality (c) follows from the definition of δ_t.

Proof of Theorem 2: Note that there exists an instance of algorithm with a threshold τ in Γ^n such that $\frac{\text{OPT}}{1+\epsilon} \leq \tau_l \leq \text{OPT}$. For this instance, it suffices to replace OPT with $\frac{\text{OPT}}{1+\epsilon}$ in the proof of Theorem 1. This proves the approximation guarantee of the theorem. For each instance of the algorithm we keep at most ℓ items. Since we have $O(\frac{\log \ell}{\epsilon})$ thresholds, the total memory complexity of the algorithm is $O(\frac{\ell \log \ell}{\epsilon})$. The update time per each element u^t for each instance is $O(km)$. This is true because we compute the gain of exchanging u^t with all the k elements of T_i^{t-1} for each function $f_i, 1 \leq i \leq m$. Therefore, the total update time per elements is $O(\frac{km \log \ell}{\epsilon})$.

D. Proof of Theorem 3

First recall that we defined:

$$S^{m,\ell} = \underset{S \subseteq \Omega, |S| \leq \ell}{\arg\max} \frac{1}{m} \sum_{i=1}^{m} \max_{|T| \leq k, T \subseteq S} f_i(T),$$

and

$$S_i^{m,\ell} = \underset{S \subseteq S^{m,\ell}, |S| \leq k}{\arg\max} f_i(S) \text{ and } \text{OPT} = \frac{1}{m} \sum_{i=1}^{m} f_i(S_i^{m,\ell}).$$

Let $\mathcal{V}(1/\mathsf{M})$ denote the distribution over random subsets of Ω where each element is picked independently with a probability $\frac{1}{\mathsf{M}}$. Define vector $\boldsymbol{p} \in [0,1]^n$ such that for $e \in \Omega$, we have

$$\boldsymbol{p}_e = \begin{cases} \mathbb{P}_{A \sim \mathcal{V}(1/\mathsf{M})}[e \in \textsc{Replacement-Greedy}(A \cup \{e\})] \text{ if } e \in S^{m,\ell}, \\ 0 \qquad \text{otherwise.} \end{cases}$$

We also define vector $\boldsymbol{p_i}$ such that for $e \in V$, we have:

$$\boldsymbol{p_i}_e = \begin{cases} \boldsymbol{p}_e \text{ if } e \in S_i^{m,\ell}, \\ 0 \qquad \text{otherwise.} \end{cases}$$

Denote by V^l the set of elements assigned to machine l. Also, let $O^l = \{e \in S^{m,\ell} : e \notin \textsc{Replacement-Greedy}(V^l \cup \{e\})\}$. Furthermore, define $O_i^l = O^l \cap S_i^{m,\ell}$. The next lemma plays a crucial role in proving the approximation guarantee of our algorithm.

Lemma 5. *Let $A \subseteq \Omega$ and $B \subseteq \Omega$ be two disjoint subsets of Ω. Suppose for each element $e \in B$, we have* $\textsc{Replacement-Greedy}(A \cup \{e\}) = \textsc{Replacement-Greedy}(A)$. *Then we have:*

$$\textsc{Replacement-Greedy}(A \cup B) = \textsc{Replacement-Greedy}(A).$$

Proof. We proof lemma by contradiction. Assume

$$\textsc{Replacement-Greedy}(A \cup B) \neq \textsc{Replacement-Greedy}(A).$$

At each iteration the element with the highest additive value is added to set S. In $\textsc{Replacement-Greedy}$, the additive value of each element depends on sets $T_i \subseteq S$. Note that sets $T_i \subseteq S$ are deterministic functions of elements of S while considering their order of additions to S. Let's assume e is the first element such that $\textsc{Replacement-Greedy}(A \cup B) \neq \textsc{Replacement-Greedy}(A)$. First note that $e \notin A$. Also, we conclude $\textsc{Replacement-Greedy}(A \cup \{e\}) \neq \textsc{Replacement-Greedy}(A)$. This contradicts with the assumption of lemma. ☐

From the definition of set O^l and Lemma 5, we have:

$$\textsc{Replacement-Greedy}(V^l) = \textsc{Replacement-Greedy}(V^l \cup O^l).$$

Lemma 6. *We have:*

$$\frac{1}{m}\sum_{i=1}^{m} f_i(T_i^l) \geq \alpha \cdot \frac{1}{m}\sum_{i=1}^{m} f_i(O_i^l),$$

where α is the approximation factor of REPLACEMENT-GREEDY.

Proof. Let OPT_i^l denote the optimum value for function f_i on the dataset $V^l \cup O^l$ for the two-stage submodular maximization problem. We have:

$$\frac{1}{m}\sum_{i=1}^{m} f_i(T_i^l) \geq \alpha \cdot \frac{1}{m}\sum_{i=1}^{m} \text{OPT}_i^l \geq \alpha \cdot \frac{1}{m}\sum_{i=1}^{m} f_i(O_i^l).$$

□

This is true because (i) REPLACEMENT-GREEDY$(V^l) =$ REPLACEMENT-GREEDY$(V^l \cup O^l)$, (ii) approximation guarantee of REPLACEMENT-GREEDY is α, and (iii) O^l and $\{O_i^l\}$ is a valid solution for the two-stage submodular maximization problem over set $V^l \cup O^l$. Assume f_i^- is the Lovász extension of a submodular function f_i.

Lemma 7 (Lemma 1, Barbosa et al. (2015))**.** *Let A be random set, and suppose that $\mathbb{E}[\mathbf{1}_A] = \lambda \cdot \boldsymbol{p}$ for a constant value of $\lambda \in [0, 1]$. Then, $\mathbb{E}[f(S)] \geq \lambda \cdot f^-(\boldsymbol{p})$.*

For each element $e \in S^{m,\ell}$ we have:

$$\begin{aligned}
\mathbb{P}[e \in O^l] &= 1 - \mathbb{P}[e \notin O^l] = 1 - \boldsymbol{p}_e, \\
\mathbb{E}[\mathbf{1}_{O^l}] &= \mathbf{1}_{S^{m,\ell}} - \boldsymbol{p}, \\
\mathbb{E}[\mathbf{1}_{O_i^l}] &= \mathbf{1}_{S_i^{m,\ell}} - \boldsymbol{p_i}.
\end{aligned}$$

Therefore, we have:

$$\mathbb{E}[\frac{1}{m}\sum_{i=1}^{m} f_i(T_i^l)] \geq \alpha \cdot \mathbb{E}[\frac{1}{m}\sum_{i=1}^{m} f_i(O_i^l)] \geq \frac{\alpha}{m} \cdot \sum_{i=1}^{m} f_i^-(\mathbf{1}_{S_i^{m,\ell}} - \boldsymbol{p_i}).$$

Furthermore, for each element $e \in S^{m,\ell}$ we have

$$\begin{aligned}
\mathbb{P}[e \in \bigcup_l S^l | e \text{ is assigned to machine } l] &= \mathbb{P}[e \in \text{REPLACEMENT-GREEDY}(V^l) | e \in V^l] \\
&= \mathbb{P}_{A\sim\mathcal{V}(1/\text{M})}[e \in \text{REPLACEMENT-GREEDY}(A) | e \in A] \\
&= \mathbb{P}_{B\sim\mathcal{V}(1/\text{M})}[e \in \text{REPLACEMENT-GREEDY}(B \cup \{e\})] \\
&= \boldsymbol{p}_e.
\end{aligned}$$

Therefore, we have

$$\mathbb{E}[\frac{1}{m}\sum_{i=1}^{m} f_j(T_i')] \geq \alpha \cdot \mathbb{E}[\frac{1}{m}\sum_{i=1}^{m} f_i(\bigcup_l S^l \cap S_i^{m,\ell})] \geq \frac{\alpha}{m} \cdot \sum_{i=1}^{m} f_i^-(\boldsymbol{p_i})$$

To Sum up above, we have:

$$\mathbb{E}[\frac{1}{m}\sum_{i=1}^{m} f_j(T_i^*)] \geq \frac{\alpha}{m}\sum_{i=1}^{m} f_j^-(\mathbf{1}_{S_i^{m,\ell}} - \boldsymbol{p_i}), \tag{13}$$

$$\mathbb{E}[\frac{1}{m}\sum_{i=1}^{m} f_i(T_i^*)] \geq \frac{\alpha}{m}\sum_{i=1}^{m} f_i^-(\boldsymbol{p_i}). \tag{14}$$

And therefore we have:

$$\mathbb{E}[\frac{1}{m}\sum_{i=1}^{m} f_i(T_i^*)] \geq \frac{\alpha}{2m}\sum_{i=1}^{m}\left[f_i^-(\boldsymbol{p_i}) + f_i^-(\mathbf{1}_{S_i^{m,\ell}} - \boldsymbol{p_i})\right] \overset{(a)}{\geq} \frac{\alpha}{2m}\sum_{i=1}^{m} f_i^-(\mathbf{1}_{S_i^{m,\ell}}) \geq \frac{\alpha}{2m}\sum_{i=1}^{m} f_i(S_i^{m,\ell}).$$

The inequality (a) results from the convexity of Lovász extensions for submodular functions. Note that the approximation guarantee of REPLACEMENT-GREEDY is $\alpha = \frac{1}{2}(1 - \frac{1}{e^2})$ (Stan et al., 2017).

E. Proof of Theorem 4

In this section, we first outline DISTRIBUTED-FAST (Algorithm 5) and then prove Theorem 4.

Algorithm 5 DISTRIBUTED-FAST

For $1 \leq l \leq \mathsf{M}$ set $V^l = \varnothing$
for $e \in \Omega$ **do**
  Assign e to a set V^l chosen uniformly at random
For $1 \leq l \leq \mathsf{M}$ sort elements of V^l based on a universal predefined ordering between elements {Any consistent ordering between elements of Ω is valid.}
Let V^l be the elements assigned to machine l
Run REPLACEMENT-PSEUDO-STREAMING on each machine l to obtain $\{S^l_\tau\}$ and $\{T^l_{\tau,i}\}$ for $1 \leq i \leq m$ and relevant values of τ on that machine
$l^*, \tau^* \leftarrow \arg\max_{l,\tau} \frac{1}{m} \sum_{i=1}^m f_i(T^l_{\tau,i})$
$S, \{T_i\} \leftarrow$ REPLACEMENT-GREEDY$(\bigcup_l \bigcup_\tau S^l_\tau)$
Return: $\arg\max\{\frac{1}{m}\sum_{i=1}^m f_i(T_i), \frac{1}{m}\sum_{i=1}^m f_i(T^{l^*}_{\tau^* i})\}$

The following lemma provides the equivalent of Lemma 5 for REPLACEMENT-PSEUDO-STREAMING. The rest of proof is exactly the same as the proof of Theorem 3 with the only difference that the approximation guarantee of REPLACEMENT-PSEUDO-STREAMING is $\gamma = \frac{1}{6+\epsilon}$.

Lemma 8. *Let $A \subseteq \Omega$ and $B \subseteq \Omega$ be two disjoint subsets of Ω. Suppose for each element $e \in B$, we have* REPLACEMENT-PSEUDO-STREAMING$(A \cup \{e\})$ = REPLACEMENT-PSEUDO-STREAMING(A). *Then we have* REPLACEMENT-PSEUDO-STREAMING$(A \cup B)$ = REPLACEMENT-PSEUDO-STREAMING(A).

Proof. First note that because of the universal predefined ordering between elements of Ω, the order of processing the elements would not change in different runs of REPLACEMENT-PSEUDO-STREAMING. Also, in the streaming setting, if an element u^t changes the set of thresholds Γ^t, then u^t would be picked by those newly instantiated thresholds. To show this, assume $\delta_{t-1} < \tau \leq \delta_t$ is one of the newly instantiated thresholds. For τ, the sets $\{T_{\tau,i}\}$ are empty and we have:

$$\tau \leq \sum_{i=1}^m \nabla_i(u^t|\varnothing) = \sum_{i=1}^m f_i(u^t) = \delta_t.$$

Therefore, u^t is added to all sets $\{T_{\tau,i}\}$. For an element $e \in B$, we have two cases: (i) e has not changed the thresholds when it is arrived, or (ii) it has instantiated new thresholds (e.g., a new threshold τ) but non of them is in the final thresholds Γ^n; because if $\tau \in \Gamma^n$, then we have $e \in S^n_\tau$, and this contradicts with the definition of set B.

Now consider REPLACEMENT-PSEUDO-STREAMING$(A \cup B)$. We prove the lemma by contradiction. Assume

$$\text{REPLACEMENT-PSEUDO-STREAMING}(A \cup B) \neq \text{REPLACEMENT-PSEUDO-STREAMING}(A).$$

Assume e is the first element of B which is picked by REPLACEMENT-PSEUDO-STREAMING$(A \cup B)$ for a threshold in Γ^n. From the above, we know that non of the thresholds Γ^n of this running instance of the algorithm is instantiated when an element of B is arrived. So, when e is arrived, all the thresholds of Γ^n which are instantiated so far are from elements of A. Also, since the order of processing of elements are fixed, REPLACEMENT-PSEUDO-STREAMING$(A \cup B)$ and REPLACEMENT-PSEUDO-STREAMING$(A \cup \{e\})$ would pick the same set of element till the point e is arrived. If e is picked by REPLACEMENT-PSEUDO-STREAMING$(A \cup B)$ for a threshold $\tau \in \Gamma^n$, then REPLACEMENT-PSEUDO-STREAMING$(A \cup \{e\})$ would also pick e for that threshold. This contradicts with the definition of set B. □

F. REPLACEMENT-GREEDY

In this section, in order to make the current manuscript self-contained, we describe the REPLACEMENT-GREEDY from (Stan et al., 2017). We use this greedy algorithm in Section 5 as one of the building blocks of our distributed algorithms.

We first define few necessary notations. The additive value of an element x to a set A from a function f_i is defined as follows:

$$\Lambda_i(x, A) = \begin{cases} f_i(x|A) & \text{if } |A| < k, \\ \max\{0, \Delta_i(x, A)\} & \text{o.w.,} \end{cases}$$

where $\Delta_i(x, A)$ is defined in Eq. (4). We also define:

$$\text{REP-GREEDY}_i(x, A) = \begin{cases} \varnothing & \text{if } |A| < k, \\ \varnothing & \Delta_i(x, A) < 0, \\ \text{REP}_i(x, A) & \text{o.w.,} \end{cases}$$

where $\text{REP}_i(x, A)$ is defined in Eq. (3). Indeed, $\text{REP-GREEDY}_i(x, A)$ represents the element from set A which should be replaced with x in order to get the maximum (positive) additive gain, where the cardinality constraint k is satisfied. REPLACEMENT-GREEDY starts with empty sets S and $\{T_i\}$. In ℓ rounds, it greedily adds elements with the maximum additive gains $\sum_{i=1}^{m} \Lambda_i(x, T_i)$ to set S. If the gain of adding these elements (or exchanging with one element of T_i where there exists k elements in T_i) is non-negative, we also update sets T_i. REPLACEMENT-GREEDY is outlined in Algorithm 6.

Algorithm 6 REPLACEMENT-GREEDY

$S \leftarrow \varnothing$ and $T_i \leftarrow \varnothing$ for all $1 \leq i \leq m$
for $1 \leq j \leq \ell$ **do**
  $x^* \geq \arg\max_{x \in \Omega} \sum_{i=1}^{m} \Lambda_i(x, T_i)$
  $S \leftarrow S + x^*$
  for $1 \leq i \leq m$ **do**
    if $\Lambda_i(x^*, T_i) > 0$ **then**
      $T_i \leftarrow T_i + x^* - \text{REP-GREEDY}_i(x^*, T_i)$
Return: S and $\{T_i\}$

G. VOC2012 Feature Explanation

To further clarify the VOC2012 dataset used in Section 6.1, we explicitly list the twenty classes that appear in the dataset. We also give an example of an image from the dataset and its corresponding characteristic vector.

0 - Aeroplane	10 - Dining Table
1 - Bicycle	11 - Dog
2 - Bird	12 - Horse
3 - Boat	13 - Motorbike
4- Bottle	14 - Person
5 - Bus	15 - Potted Plant
6 - Car	16 - Sheep
7 - Cat	17 - Sofa
8 - Chair	18 - Train
9 - Cow	19 - TV Monitor

(a)

(b)

Figure 5. (a) shows the twenty classes that appear in the *VOC2012* dataset. The number adjacent to each class represents the index of that class in the characteristic vector associated with each image. For example, the image shown in (b) contains one boat, one bird, and one person. Therefore, the characteristic vector for this image is [0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]. This also means that the image in (b) appears in the sets Ω_2, Ω_4, and Ω_{14}.