

# Robust Subset Selection by Greedy and Evolutionary Pareto Optimization

Chao Bian, Yawen Zhou and Chao Qian\*

State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China  
bianc@lamda.nju.edu.cn, zyw8769@gmail.com, qianc@lamda.nju.edu.cn

## Abstract

Subset selection, which aims to select a subset from a ground set to maximize some objective function, arises in various applications such as influence maximization and sensor placement. In real-world scenarios, however, one often needs to find a subset which is robust against (i.e., is good over) a number of possible objective functions due to uncertainty, resulting in the problem of robust subset selection. This paper considers robust subset selection with monotone objective functions, relaxing the submodular property required by previous studies. We first show that the greedy algorithm can obtain an approximation ratio with respect to the correlation and submodularity ratios of the objective functions; and then propose EPORSS, an evolutionary Pareto optimization algorithm that can utilize more time to find better subsets. We prove that EPORSS can also be theoretically grounded, achieving a similar approximation guarantee to the greedy algorithm. In addition, we derive the lower bound of the correlation ratio for the application of robust influence maximization, and further conduct experiments to validate the performance of the greedy algorithm and EPORSS.

## 1 Introduction

The subset selection problem is to select a subset of size at most  $k$  from a total set of  $n$  items for optimizing some objective function  $f$ , which arises in many applications, such as influence maximization [Kempe *et al.*, 2003], sensor placement [Krause *et al.*, 2008b], document summarization [Lin and Bilmes, 2011], and unsupervised feature selection [Feng *et al.*, 2019], etc. This problem is NP-hard in general, and much effort has been devoted to the design of polynomial-time approximation algorithms.

The greedy algorithm, which iteratively selects one item with the largest marginal gain on  $f$ , is a very popular algorithm for solving the subset selection problem. For a

\*This work was supported by the NSFC (62022039) and the CAAI-Huawei MindSpore Open Fund. Chao Qian is the corresponding author. The supplementary material is available at <https://arxiv.org/abs/2205.01415>.

monotone submodular objective function  $f$ , it can achieve the optimal polynomial-time approximation guarantee of  $1 - 1/e$  [Nemhauser and Wolsey, 1978]. Due to its high-efficiency and theoretically-grounded performance, it is most favored and has been studied extensively [Minoux, 1978; Mirzasoleiman *et al.*, 2015; Hassidim and Singer, 2017; Qian *et al.*, 2018; Friedrich *et al.*, 2019].

In real-world applications of subset selection, however, we may want to find a subset which is robust against a number of possible objective functions, i.e., the subset is good on the worst of these objective functions. For example, in influence maximization, the influence of a set of social network users has significant uncertainty due to different models and parameters, and the goal is to optimize a set of influence functions simultaneously, in which one function is assured to describe the influence process exactly (but which one is not told) [He and Kempe, 2016]. In sensor placement, sensors are often used to monitor various phenomena such as temperature and humidity at the same time, whose observations are modeled by different functions, and the goal is to find a subset of candidate locations to place sensors such that the sensors can perform well on all the objective functions [Anari *et al.*, 2019].

To address the above more complicated real-world scenario of subset selection, Krause *et al.* [2008a] considered the robust subset selection problem

$$\arg \max_{X \subseteq V} \min_{1 \leq i \leq m} f_i(X) \quad \text{s.t.} \quad |X| \leq k, \quad (1)$$

where  $f_1, f_2, \dots, f_m$  are monotone submodular objective functions and  $|\cdot|$  denotes the size of a set. They proved that there cannot exist any polynomial time approximation algorithm unless  $P=NP$ , and proposed an algorithm SATURATE, which can find a subset matching the optimal worst-case objective value but with size larger than  $k$ . Anari *et al.* [2019] extended the cardinality constraint  $|X| \leq k$  in Eq. (1) to a matroid constraint, and provided a bi-criteria algorithm which can find  $O(\log(m/\epsilon))$  feasible subsets whose union achieves an approximation ratio of  $1 - \epsilon$ , where  $\epsilon \in (0, 1)$ . Ud-wani [2018] designed a fast and practical algorithm for the case of  $m = o(k)$ . He and Kempe [2016] modified SATURATE for the application of robust influence maximization, and showed that a  $(1 - 1/e)$ -approximation ratio can be achieved when enough extra seed users may be selected. Iyer [2019] considered both robust submodular minimization and maximization problems, and proposed a general framework to transform the algorithm for one problem to the other.

Recently, Hou and Clark [2021] introduced the notion of correlation ratio and proposed a modified greedy algorithm which can achieve an approximation guarantee with respect to the correlation ratio.

Note that all the above-mentioned previous studies considered monotone submodular objective functions. However, the objective functions of some applications such as influence maximization under general cascade model [Kempe *et al.*, 2003] and sparse regression [Das and Kempe, 2011] can be non-submodular. This paper thus considers the problem of robust subset selection with monotone objective functions, that is,  $f_1, f_2, \dots, f_m$  are monotone but not necessarily submodular. First, we show that the classic greedy algorithm which iteratively adds one item maximizing the gain on the worst-case objective function can achieve an approximation ratio of  $1 - e^{-\beta\gamma}$ , where  $\beta$  and  $\gamma$  are the correlation ratio (slightly different from that proposed in [Hou and Clark, 2021]) and the submodularity ratio of the objective functions, respectively.

Though the greedy algorithm is efficient and theoretically grounded, its performance may be limited in practice due to the greedy nature. Next, we propose an evolutionary Pareto optimization algorithm, EPORSS, which is an anytime algorithm that can use more time to find better subsets. EPORSS first reformulates the original robust subset selection problem as a bi-objective optimization problem that maximizes  $\min_{1 \leq i \leq m} f_i$  and minimizes the subset size simultaneously, then employs a multi-objective evolutionary algorithm to solve it, and finally selects the best feasible solution (i.e., subset) from the generated population. We prove that EPORSS can achieve an approximation ratio similar to that of the greedy algorithm.

The submodularity ratio  $\gamma$  has been analyzed before (e.g., [Bian *et al.*, 2017; Elenberg *et al.*, 2018; Qian *et al.*, 2019]), and here we prove a lower bound of the correlation ratio  $\beta$  for the application of robust influence maximization, to show the applicability of the derived approximation guarantees of the greedy algorithm and EPORSS. Experimental results on robust influence maximization show that the greedy algorithm can achieve competitive performance to the well-known algorithm SATURATE, but requires less time; EPORSS using more time can achieve superior performance over the greedy algorithm and SATURATE. Therefore, these two algorithms can be used under the limited and abundant computation budgets, respectively, helping solve diverse practical problems of robust subset selection better.

## 2 Preliminaries

Given a finite nonempty set  $V = \{v_1, v_2, \dots, v_n\}$ , we study the functions  $f : 2^V \rightarrow \mathbb{R}$  defined on subsets of  $V$ . A set function  $f$  is monotone if for any  $X \subseteq Y$ ,  $f(X) \leq f(Y)$ . In this paper, we consider monotone functions and assume that they are normalized, i.e.,  $f(\emptyset) = 0$ . A set function  $f$  is submodular [Nemhauser *et al.*, 1978] if for any  $X \subseteq Y \subseteq V$ ,

$$f(Y) - f(X) \leq \sum_{v \in Y \setminus X} (f(X \cup \{v\}) - f(X)).$$

The submodularity ratio in Definition 1 characterizes how close a set function  $f$  is to submodularity. For a monotone

function  $f$ , it holds that  $0 \leq \gamma_{X,b}(f) \leq 1$ , and  $f$  is submodular iff  $\gamma_{X,b}(f) = 1$  for any  $X$  and  $b$ . For some concrete non-submodular applications, lower bounds on  $\gamma_{X,b}(f)$  have been derived [Bian *et al.*, 2017; Elenberg *et al.*, 2018; Qian *et al.*, 2019]. When  $f$  is clear, we will omit  $f$  and use  $\gamma_{X,b}$  for short.

**Definition 1** (Submodularity Ratio [Das and Kempe, 2011]). *Let  $f$  be a non-negative set function. The submodularity ratio of  $f$  with respect to a set  $X$  and a parameter  $b \geq 1$  is*

$$\gamma_{X,b}(f) = \min_{L \subseteq X, S: |S| \leq b, S \cap L = \emptyset} \frac{\sum_{v \in S} (f(L \cup \{v\}) - f(L))}{f(L \cup S) - f(L)}.$$

The subset selection problem as presented in Definition 2 is to select a subset  $X$  of  $V$  such that a given monotone objective function  $f$  is maximized with the constraint  $|X| \leq k$ , where  $|\cdot|$  denotes the size of a set.

**Definition 2** (Subset Selection). *Given a set of items  $V = \{v_1, v_2, \dots, v_n\}$ , a monotone objective function  $f$ , and a budget  $k$ , to find*

$$\arg \max_{X \subseteq V} f(X) \quad \text{s.t.} \quad |X| \leq k.$$

It has many applications such as influence maximization [Kempe *et al.*, 2003] and sensor placement [Krause *et al.*, 2008b]. Influence maximization is to select a subset of users from a social network to maximize its influence spread, which often appears in viral marketing. Sensor placement is to select a few places to install sensors such that the uncertainty is mostly reduced, which often appears in spatial monitoring of certain phenomena.

The original subset selection problem in Definition 2 considers optimizing only one objective function. However, in real-world scenarios, one may need to find a subset which performs well on a number of possible objective functions. For example, in viral marketing, the company may want to identify a set of users to promote several products (e.g., shirts and sweaters) simultaneously, while the influences of users on different products can be different [Kalimeris *et al.*, 2019]; even for one product, perturbations or uncertainty regarding the parameters of diffusion models will result in different objective functions. In spatial monitoring of certain phenomena, sensors are often used to measure various parameters such as temperature and humidity at the same time, giving rise to different objective functions [Anari *et al.*, 2019]; besides, the data obtained in real world is often noisy, which will lead to perturbations on the objective functions. In both above-mentioned applications, we need to find a subset that is robust against all the possible objective functions, i.e., optimize the worst of these objective functions. More evidences can be seen on the applications of robust column subset selection [Zhu *et al.*, 2015] and robust sparse regression [Thompson, 2022]. In this paper, we focus on the general robust subset selection problem, presented as follows.

**Definition 3** (Robust Subset Selection). *Given a set of items  $V = \{v_1, v_2, \dots, v_n\}$ ,  $m$  monotone objective functions  $f_1, f_2, \dots, f_m$ , and a budget  $k$ , to find*

$$\arg \max_{X \subseteq V} F(X) \quad \text{s.t.} \quad |X| \leq k,$$

where  $F(X) = \min_{1 \leq i \leq m} f_i(X)$  denotes the worst-case objective value of  $X$ .

This problem is very hard even when the objective functions  $f_1, f_2, \dots, f_m$  satisfy the submodular property, because it has been proved that polynomial-time algorithms cannot achieve any constant approximation ratio. By introducing a quantity  $\beta^*$ , which reflects the correlation of the objective functions, Hou and Clark [2021] proposed the modified greedy algorithm with an approximation ratio of  $1 - e^{-\beta^*}$ . If allowing the subset size larger than  $k$ , Krause *et al.* [2008a] showed that the SATURATE algorithm can achieve the optimal  $F$  value.

Note that the correlation of the objective functions relies on the behavior of the modified greedy algorithm [Hou and Clark, 2021]. Here we generalize this notion in Definition 4, which will be used in our theoretical analysis. Intuitively, the correlation ratio reflects the similarity of the objective functions to be optimized. A large correlation ratio implies that all the objective functions can be improved substantially by adding a common item into the subset  $X$ . For monotone objective functions,  $0 \leq \beta_X \leq 1$ , and  $\beta_X = 1$  means that there exists one common item, adding which makes all the functions achieve the largest improvement.

**Definition 4** (Correlation Ratio). *The correlation ratio of functions  $f_1, f_2, \dots, f_m$  with respect to a set  $X$  is*

$$\beta_X = \max_{v \in V \setminus X} \min_{1 \leq i \leq m} \frac{f_i(X \cup \{v\}) - f_i(X)}{f_i(X \cup \{v^i\}) - f_i(X)},$$

where  $v^i \in \arg \max_{v \in V \setminus X} f_i(X \cup \{v\})$ .

### 3 The Greedy Algorithm

To solve the robust subset selection problem, a nature way is to apply the greedy process. As shown in Algorithm 1, the greedy algorithm iteratively adds one item with the largest improvement on  $F$  until  $k$  items are selected. Let  $\text{OPT} = \max_{X \subseteq V, |X| \leq k} F(X)$  denote the optimal function value of the problem in Definition 3, and  $X_j$  ( $0 \leq j \leq k$ ) denote the subset generated in the  $j$ -th iteration (note that  $X_0 = \emptyset$ ) of Algorithm 1. We prove in Theorem 1 that the greedy algorithm can achieve an approximation ratio of  $1 - e^{-\beta\gamma}$ , where  $\beta = \min_{0 \leq j \leq k-1} \beta_{X_j}$  denotes the minimum correlation ratio with respect to the first  $k$  subsets generated by the greedy algorithm, and  $\gamma = \min_{1 \leq i \leq m} \gamma_{X_{k-1}, k}(f_i)$  denotes the minimum submodularity ratio of all the objective functions with respect to  $X_{k-1}$  and  $k$ .

**Theorem 1.** *For robust subset selection in Definition 3, the greedy algorithm finds a subset  $X \subseteq V$  with  $|X| = k$  and*

$$F(X) \geq (1 - e^{-\beta\gamma}) \cdot \text{OPT}.$$

When the number  $m$  of functions is 1, i.e., the robust subset selection problem in Definition 3 degrades to the original subset selection problem in Definition 2, we have  $\beta = 1$ , and the approximation ratio becomes  $1 - e^{-\gamma}$ , which is the optimal polynomial-time approximation guarantee [Harshaw *et al.*, 2019]. When  $f_1, f_2, \dots, f_m$  are submodular, we have  $\gamma = 1$ , and the approximation ratio becomes  $1 - e^{-\beta}$ , which is similar to that achieved by the modified greedy algorithm [Hou and Clark, 2021]. Though the two approximation guarantees cannot be compared exactly due to the slightly

---

#### Algorithm 1 Greedy Algorithm

---

**Input:** all items  $V = \{v_1, v_2, \dots, v_n\}$ , the objective function  $F = \min_{1 \leq i \leq m} f_i$ , and a budget  $k$

**Output:** a subset of  $V$  with  $k$  items

**Process:**

- 1: Let  $j = 0$  and  $X_j = \emptyset$ ;
  - 2: **while**  $j < k$  **do**
  - 3:   Let  $v^* = \arg \max_{v \in V \setminus X_j} F(X_j \cup \{v\})$ ;
  - 4:   Let  $X_{j+1} = X_j \cup \{v^*\}$ , and  $j = j + 1$
  - 5: **end while**
  - 6: **return**  $X_k$
- 

different definitions of the correlation ratio, our greedy algorithm is faster than the modified greedy algorithm [Hou and Clark, 2021]. The greedy algorithm in Algorithm 1 performs  $(n - k/2 + 1/2)k$  number of worst-case objective function evaluations in total, while the modified greedy algorithm needs to first compute the largest improvement on each objective function in each iteration, thus requires extra  $(n - k/2 + 1/2)k$  number of worst-case objective function evaluations in total. The detailed procedure of the modified greedy algorithm is provided in the supplementary material due to space limitation.

The proof of Theorem 1 relies on Lemma 2, which intuitively means that for any subset, the inclusion of one specific item can improve  $F$  by at least a quantity proportional to the current distance to the optimum. Lemma 1 shows a similar result with respect to a single objective function  $f_i$ , which will be used in the proof of Lemma 2. The detailed proof of Lemma 2 is provided in the supplementary material due to space limitation.

**Lemma 1.** [Qian *et al.*, 2016] *For any  $1 \leq i \leq m$  and  $X \subseteq V$ , there exists one item  $v \in V \setminus X$  such that*

$$f_i(X \cup \{v\}) - f_i(X) \geq (\gamma_{X, k}(f_i)/k) \cdot (\text{OPT}_i - f_i(X)).$$

where  $\text{OPT}_i = \max_{X \subseteq V, |X| \leq k} f_i(X)$  denotes the optimal function value of the problem in Definition 2 with respect to the objective function  $f_i$ .

**Lemma 2.** *For any  $X \subseteq V$ , there exists one item  $v \in V \setminus X$  such that*

$$F(X \cup \{v\}) - F(X) \geq (\beta_X \gamma_{X, k}^{\min}/k) \cdot (\text{OPT} - F(X)),$$

where  $\gamma_{X, k}^{\min} = \min_{1 \leq i \leq m} \gamma_{X, k}(f_i)$  denotes the minimum submodularity ratio of all the objective functions with respect to  $X$  and  $k$ .

*Proof of Theorem 1.* We will show that for any  $0 \leq j \leq k$ ,

$$\text{OPT} - F(X_j) \leq (1 - \beta\gamma/k)^j \cdot \text{OPT}. \quad (2)$$

For  $j = 0$ , Eq. (2) holds because  $X_0 = \emptyset$  and  $\forall i : f_i(\emptyset) = 0$ . Suppose that Eq. (2) holds for some  $j$  (where  $0 \leq j \leq k - 1$ ), we will prove that it also holds for  $j + 1$ . Let  $v^* \in \arg \max_{v \in V \setminus X_j} F(X_j \cup \{v\})$ . By Lemma 2, we have

$$\begin{aligned} F(X_j \cup \{v^*\}) - F(X_j) &\geq (\beta_{X_j} \gamma_{X_j, k}^{\min}/k) \cdot (\text{OPT} - F(X_j)) \\ &\geq (\beta\gamma/k) \cdot (\text{OPT} - F(X_j)), \end{aligned}$$

where the second inequality is by  $\beta = \min_{0 \leq j \leq k-1} \beta_{X_j} \leq \beta_{X_j}$ , and  $\gamma = \min_{1 \leq i \leq m} \gamma_{X_{k-1}, k}(f_i) \leq \min_{1 \leq i \leq m} \gamma_{X_j, k}(f_i) = \gamma_{X_j, k}^{\min}$  due to the fact that  $\gamma_{X, b}(f) \leq \gamma_{Y, b}(f)$  for any  $Y \subseteq X$ . According to lines 3-4 in Algorithm 1,  $X_{j+1}$  is generated by combining  $X_j$  with  $v^*$ . Thus, we have

$$\begin{aligned} \text{OPT} - F(X_{j+1}) &= \text{OPT} - F(X_j \cup \{v^*\}) \\ &\leq (1 - \beta\gamma/k)(\text{OPT} - F(X_j)) \leq (1 - \beta\gamma/k)^{j+1} \text{OPT}, \end{aligned}$$

where the last inequality holds by the induction hypothesis.

Thus, we have shown that Eq. (2) holds for any  $0 \leq j \leq k$ . By setting  $j = k$  and applying  $1 + x \leq e^x$  for any  $x \in \mathbb{R}$ , the theorem holds.  $\square$

## 4 The EPORSS Algorithm

Though the greedy algorithm is efficient and its approximation performance can be theoretically bounded, its performance may be limited in practice due to the greedy nature. Thus, we further propose an Evolutionary Pareto Optimization [Friedrich and Neumann, 2015; Qian *et al.*, 2015] algorithm for Robust Subset Selection, called EPORSS, which can use more time to find better solutions.

Let a Boolean vector  $\mathbf{x} \in \{0, 1\}^n$  represent a subset  $X$  of  $V$ , where the  $i$ -th bit  $x_i = 1$  if  $v_i \in X$  and  $x_i = 0$  otherwise. For convenience, we will not distinguish  $\mathbf{x}$  and its corresponding subset. EPORSS reformulates the original problem in Definition 3 as a bi-objective maximization problem:

$$\begin{aligned} &\arg \max_{\mathbf{x} \in \{0, 1\}^n} (g_1(\mathbf{x}), g_2(\mathbf{x})), \\ \text{where } g_1(\mathbf{x}) &= \begin{cases} -\infty, & |\mathbf{x}| \geq 2k \\ F(\mathbf{x}), & \text{otherwise} \end{cases}, \quad g_2(\mathbf{x}) = -|\mathbf{x}|. \end{aligned}$$

That is, EPORSS maximizes the original objective  $F$  and minimizes the subset size  $|\mathbf{x}|$  simultaneously. Note that infeasible solutions can cause “shortcuts” and be useful in the evolutionary search, thus  $g_1$  only punishes the solutions with size at least  $2k$  so that the infeasible solutions with size belonging to  $\{k+1, \dots, 2k-1\}$  can participate in the optimization process. Meanwhile, too many infeasible solutions may lead to longer runtime, thus the setting of the threshold  $2k$  is to make a tradeoff between performance and runtime.

In the bi-objective setting, the domination relationship in Definition 5 is used to compare two solutions. Note that  $\mathbf{x}$  and  $\mathbf{y}$  are *incomparable* if neither  $\mathbf{x} \succeq \mathbf{y}$  nor  $\mathbf{y} \succeq \mathbf{x}$ .

**Definition 5** (Domination). *For two solutions  $\mathbf{x}$  and  $\mathbf{y}$ ,*

- $\mathbf{x}$  weakly dominates  $\mathbf{y}$  (denoted as  $\mathbf{x} \succeq \mathbf{y}$ ) if  $g_1(\mathbf{x}) \geq g_1(\mathbf{y}) \wedge g_2(\mathbf{x}) \geq g_2(\mathbf{y})$ ;
- $\mathbf{x}$  dominates  $\mathbf{y}$  (denoted as  $\mathbf{x} \succ \mathbf{y}$ ) if  $\mathbf{x} \succeq \mathbf{y}$  and  $g_1(\mathbf{x}) > g_1(\mathbf{y}) \vee g_2(\mathbf{x}) > g_2(\mathbf{y})$ .

EPORSS as described in Algorithm 2 employs a simple multi-objective evolutionary algorithm [Friedrich and Neumann, 2015; Qian *et al.*, 2019] to optimize the bi-objective problem. It starts from the empty set  $\{0\}^n$  (line 1). In each iteration, a new solution  $\mathbf{x}'$  is generated by applying the bit-wise mutation operator to an archived solution  $\mathbf{x}$  randomly selected from the current population  $P$  (lines 3–4); if  $\mathbf{x}'$  is not dominated by any previously archived solution (line 5), it will be added into  $P$ , and meanwhile those solutions weakly

---

## Algorithm 2 EPORSS Algorithm

---

**Input:** all items  $V = \{v_1, \dots, v_n\}$ , the objective function  $F = \min_{1 \leq i \leq m} f_i$ , and a budget  $k$

**Parameter:** the number  $T$  of iterations

**Output:** a subset of  $V$  with at most  $k$  items

**Process:**

- 1: Let  $\mathbf{x} = \{0\}^n$ ,  $P = \{\mathbf{x}\}$ , and let  $t = 0$ ;
  - 2: **while**  $t < T$  **do**
  - 3:   Select  $\mathbf{x}$  from  $P$  uniformly at random;
  - 4:   Generate  $\mathbf{x}'$  by flipping each bit of  $\mathbf{x}$  with prob.  $1/n$ ;
  - 5:   **if**  $\nexists z \in P$  such that  $z \succ \mathbf{x}'$  **then**
  - 6:      $P = (P \setminus \{z \in P \mid \mathbf{x}' \succeq z\}) \cup \{\mathbf{x}'\}$
  - 7:   **end if**
  - 8:    $t = t + 1$
  - 9: **end while**
  - 10: **return**  $\arg \max_{\mathbf{x} \in P, |\mathbf{x}| \leq k} F(\mathbf{x})$
- 

dominated by  $\mathbf{x}'$  will be removed (line 6). After running  $T$  iterations, the solution with the largest  $F$  value satisfying the size constraint in  $P$  is selected (line 10).

We can see from the algorithm procedure that EPORSS naturally maintains a population of non-dominated solutions due to the bi-objective transformation. Meanwhile, the bit-wise mutation operator for reproduction is a global search operator, which flips each bit of a solution independently with probability  $1/n$ . These characteristics may make EPORSS have a better ability of escaping from local optima than the greedy algorithm. Our experiments will show that EPORSS can find better solutions by using more time. Here, we first show that though using an evolutionary process for optimization, EPORSS can also be theoretically grounded. Particularly, Theorem 2 shows that EPORSS can achieve a similar approximation guarantee as the greedy algorithm. Note that  $\mathbb{E}[T]$  denotes the expected number  $T$  of iterations,  $\gamma' = \min_{1 \leq i \leq m} \min_{X: |X|=k-1} \gamma_{X, k}(f_i)$ , and  $\beta' = \min_{X: |X| \leq k-1} \beta_X$ .

**Theorem 2.** *For robust subset selection in Definition 3, EPORSS with  $\mathbb{E}[T] \leq 2ek^2n$  finds a subset  $X \subseteq V$  with  $|X| \leq k$  and  $F(X) \geq (1 - e^{-\beta' \gamma'}) \cdot \text{OPT}$ .*

The proof can be accomplished by following the proof of Theorem 2 in [Qian *et al.*, 2019] and using Lemma 2. The main idea is analyzing the increase of a quantity  $J_{\max}$ , which is defined as the largest number  $j \in \{0, 1, \dots, k\}$  such that in the population  $P$ ,  $\exists \mathbf{x}$  with  $|\mathbf{x}| \leq j$  and  $F(\mathbf{x}) \geq (1 - (1 - \beta' \gamma'/k)^j) \cdot \text{OPT}$ . For the sake of completeness, we still provide the full proof in the supplementary material.

## 5 Analysis of the Correlation Ratio

In the above two sections, we have shown that the approximation performance of the greedy algorithm and EPORSS can be theoretically bounded. However, the approximation guarantees depend on the submodularity ratio  $\gamma_{X, b}$  in Definition 1 and the correlation ratio  $\beta_X$  in Definition 4. Previous studies have shown that the lower bound on  $\gamma_{X, b}$  can be derived for several non-submodular objective functions [Bian *et al.*, 2017; Elenberg *et al.*, 2018]. In this section, we further

derive a lower bound on  $\beta_X$  for the application of robust influence maximization, showing the applicability of the derived approximation guarantees.

Influence maximization is to identify a set of influential users in social networks. Let a directed graph  $G(V, E)$  represent a social network, where each node represents a user and each edge  $(u, v)$  has a probability  $p_{u,v}$  representing the strength of influence from user  $u$  to user  $v$ . As presented in Definition 6, the goal is to find a subset  $X$  with size at most  $k$  such that the expected number (denoted as  $\sigma(X)$ ) of nodes activated by propagating from  $X$  is maximized.

**Definition 6** (Influence Maximization). *Given a directed graph  $G(V, E)$ , edge probabilities  $p_{u,v}$  where  $(u, v) \in E$  is an edge, and a budget  $k$ , to find*

$$\arg \max_{X \subseteq V} \sigma(X) \quad \text{s.t.} \quad |X| \leq k.$$

Independence Cascade (IC) is a fundamental propagation model [Kempe *et al.*, 2003] to estimate the influence function  $\sigma$ , and will also be used in our analysis. Starting from a seed set  $X$ , it uses a set  $A_t$  to record the nodes activated at time  $t$ , and at time  $t + 1$ , each inactive neighbour  $v$  of  $u \in A_t$  becomes active with probability  $p_{u,v}$ . This process is repeated until no nodes get activated at some time. Under the IC model, the objective function  $\sigma$  is monotone submodular [Kempe *et al.*, 2003].

For robust influence maximization, we are given several probability vectors  $\theta^1, \theta^2, \dots, \theta^m$ , where each  $\theta$  corresponds to a set of edge probabilities, and thus determines a specific influence function, denoted as  $\sigma_\theta$ . The goal is to maximize the worst of the  $m$  influence functions  $\sigma_{\theta^1}, \sigma_{\theta^2}, \dots, \sigma_{\theta^m}$ , as shown in the following definition.

**Definition 7** (Robust Influence Maximization [He and Kempe, 2016; Kalimeris *et al.*, 2019]). *Given a directed graph  $G(V, E)$ ,  $m$  probability vectors  $\theta^1, \theta^2, \dots, \theta^m$ , and a budget  $k$ , to find*

$$\arg \max_{X \subseteq V} \min_{1 \leq i \leq m} \sigma_{\theta^i}(X) \quad \text{s.t.} \quad |X| \leq k.$$

Theorem 3 gives a lower bound on  $\beta_X$  for this application. The proof relies on Lemma 3, which intuitively means that when two probability vectors are close, the difference of the influence spread determined by these two vectors can also be small. Lemma 3 is inspired from [Chen *et al.*, 2016], but slightly refined. The detailed proof is provided in the supplementary material due to space limitation, and we introduce the main proof idea here. Under the IC model, the influence spread  $\sigma_\theta(X)$  can be calculated as  $\sum_{S \subseteq G} \pi_\theta(S) \cdot \sigma_S(X)$  [Kempe *et al.*, 2003], where  $\pi_\theta(S)$  denotes the probability of sampling a subgraph  $S$  from  $G$  according to the edge probability vector  $\theta$ , that is, each edge  $(u, v) \in E$  appears in  $S$  with probability  $p_{u,v}$ ; and  $\sigma_S(X)$  denotes the number of nodes that can be reachable from  $X$  on the given subgraph  $S$ , which is deterministic when  $S$  is given. When two probability vectors only have little difference, the probabilities of sampling a specific subgraph are also slightly different, leading to the similar influence spread.

**Lemma 3.** *For any two probability vectors  $\theta$  and  $\theta'$ , let  $\delta(\theta, \theta') = \sum_i |\theta_i - \theta'_i|$  denote the difference between the two vectors, where  $\theta_i$  and  $\theta'_i$  denote the  $i$ -th element of  $\theta$  and*

*$\theta'$ , respectively, and  $|\cdot|$  denotes the absolute value of a real number. Then, for any  $X \subseteq V$ , we have  $|\sigma_\theta(X) - \sigma_{\theta'}(X)| \leq n \cdot \delta(\theta, \theta')$ , where  $n$  denotes the size of  $V$ .*

Theorem 3 shows that  $\beta_X$  can be lower bounded by  $1 - 2en \cdot \delta_{\max}$ . This is intuitive, because  $\delta_{\max}$  measures the difference among the probability vectors, and thus the corresponding objective functions, while  $\beta_X$  characterizes the similarity among the objective functions. Thus, a smaller  $\delta_{\max}$  will lead to a larger lower bound on  $\beta_X$ . The proof is accomplished by analyzing the ratio of marginal gains on  $\sigma_{\theta^i}$  by respectively adding  $v^1$  and  $v^i$  to  $X$ , where  $v^1$  and  $v^i$  are the items whose inclusion in  $X$  can make  $\sigma_{\theta^1}$  and  $\sigma_{\theta^i}$  achieve the largest improvement, respectively. The detailed proof is provided in the supplementary material due to space limitation.

**Theorem 3.** *For robust influence maximization in Definition 7, let  $\delta_{\max} = \max_{1 \leq i, j \leq m, i \neq j} \delta(\theta^i, \theta^j)$  denote the maximum difference among the  $m$  probability vectors. Then, for any  $X \in \{X_0, X_1, \dots, X_{k-1}\}$ , we have  $\beta_X \geq 1 - 2en \cdot \delta_{\max}$ , where  $X_j$  ( $0 \leq j \leq k - 1$ ) denotes the subset generated in the  $j$ -th iteration (note that  $X_0 = \emptyset$ ) of Algorithm 1.*

## 6 Experiments

In this section, we empirically compare the performance of the greedy algorithm, EPORSS and two previous algorithms, modified greedy [Hou and Clark, 2021] and SATURATE [Krause *et al.*, 2008a], on the application of robust influence maximization. As suggested in [Krause *et al.*, 2008a], we set the parameter  $\alpha$  of SATURATE to 1 such that the output subset will not violate the cardinality constraint. The number of iterations of EPORSS is set to  $\lfloor 2ek^2n \rfloor$  as suggested by Theorem 2. As EPORSS is randomized, we repeat its run for 10 times independently, and report the average results and standard deviation.

The experiments are performed on two real-world data sets, *ego-Facebook* and *as-733*, downloaded from <https://snap.stanford.edu/data/index.html>. The *ego-Facebook* data set contains one network which consists of friend links from Facebook, and we use the perturbation interval method [He and Kempe, 2016] to sample multiple networks with different edge probabilities (which give rise to different objective functions). That is, we first estimate the propagation probability of one edge from node  $u$  to  $v$  by  $p_{u,v} = \frac{\text{weight}(u,v)}{\text{indegree}(v)}$ , as widely used in [Chen *et al.*, 2009; Goyal *et al.*, 2011]; and then sample from the interval  $[0.9 \cdot p_{u,v}, 1.1 \cdot p_{u,v}]$  multiple times to obtain different probabilities. For estimating the influence spread, we use the IC model as presented in Section 5, which results in monotone submodular objective functions.

The *as-733* data set contains 733 communication networks spanning an interval of 785 days, where the nodes and edges may be added or deleted over time. Thus, the two data sets reflect different causes for uncertainty: the edge probabilities are subject to perturbations, and the influence functions are learned from the networks with different structures. We use the general IC model [Kempe *et al.*, 2003] to estimate the influence spread on *as-733*. For the general IC model, the probability of activating  $v$  by  $u$  is  $p_v(u, S)$  instead of  $p_{u,v}$ , where  $S$  is the set of neighbours that have already tried (and

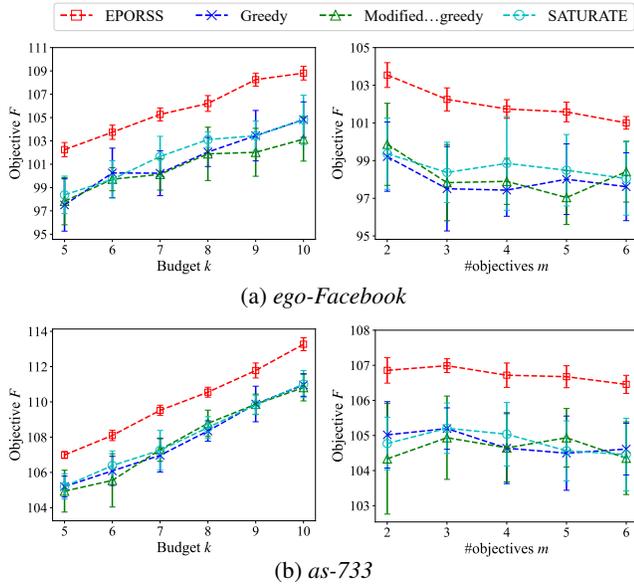


Figure 1: Robust influence maximization on *ego-Facebook* and *as-733*. The objective  $F$ : the worst-case influence spread (the larger, the better). The left subfigure: objective  $F$  vs. budget  $k$ ; the right subfigure: objective  $F$  vs. #objectives  $m$ .

failed) to activate  $v$ . Specifically, for each network in *as-733*, we set the probability  $p_v(u, S)$  to  $\min\{0.1 + 0.05 \cdot |S|, 1\}$ , i.e., the probability of activating  $v$  is 0.1 for the first try, and then the probability increases by 0.05 once a try fails. Note that under the general IC model, the objective functions are generally non-submodular.

For both data sets, we examine the performance of the compared algorithms when the budget  $k$  changes from 5 to 10 (where the number  $m$  of objective functions is fixed to 3), and also consider the scenario where the number  $m$  of objective functions changes from 2 to 6 (where the budget  $k$  is fixed to 5). Note that we sample the most active 200 nodes from these two data sets for examination, as in [He and Kempe, 2016]. To estimate the influence spread  $\sigma(X)$  of a subset  $X$  of nodes, we simulate the diffusion process 100 times independently and use the average as an estimation. As the objective evaluation is noisy, we also repeat the running of the three deterministic algorithms (i.e., the greedy algorithm, the modified greedy algorithm and SATURATE) 10 times independently and report the average results. We can observe from Figure 1 that the three deterministic algorithms achieve the similar performance, while EPORSS is much better.

**Running time comparison.** The greedy algorithm in algorithm 1 needs to perform  $(n - j + 1)$  number of worst-case objective function evaluations in the  $j$ -th iteration. Thus, the total running time, i.e., the number of worst-case objective function evaluations, is  $\sum_{j=1}^k (n - j + 1) = (n - k/2 + 1/2)k$ . The modified greedy algorithm starts from an empty set, and iteratively adds one item  $v$  into the current subset  $X$ , such that all the objective functions can be improved as much as possible. However, for each of the  $m$  objective functions, the improvement is measured by the ratio of the marginal gain achieved by adding  $v$  and the largest possible marginal

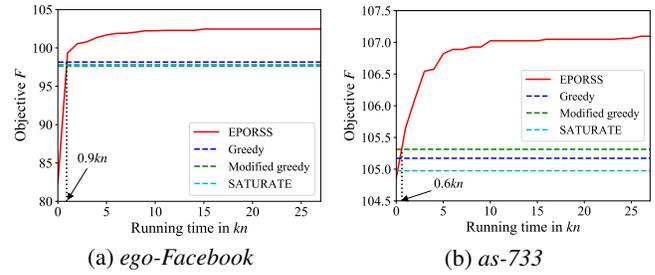


Figure 2: The objective  $F$  vs. the running time on robust influence maximization with  $m = 3$  and  $k = 5$ . The objective  $F$ : the worst-case influence spread (the larger, the better).

gain, which needs to be computed in the first place and thus requires extra  $(n - j + 1)$  number of objective function evaluations in the  $j$ -th iteration. The process terminates after  $k$  items are selected. Thus, the total running time is  $\sum_{j=1}^k 2(n - j + 1) = (2n - k + 1)k$ , i.e., double that of the greedy algorithm. Though the extra objective function evaluations can be avoided by storing the marginal gain achieved by adding each item, extra storage cost is required. SATURATE employs a binary search framework, and each shrink over the search interval needs to perform a greedy subroutine, implying that the running time of SATURATE is substantially larger than that of the greedy algorithm. Thus, the greedy algorithm can achieve competitive performance to the modified greedy algorithm and SATURATE using less time.

EPORSS can achieve the best performance using more time, i.e.,  $2ek^2n$  iterations. However,  $2ek^2n$  is only a theoretical upper bound, which may be too pessimistic in practice. By selecting the three deterministic algorithms (i.e., the greedy algorithm, the modified greedy algorithm and SATURATE) as the baselines, we plot the curve of the objective  $F$  over the running time of EPORSS on the two data sets, as shown in Figure 2. The  $x$ -axis is in  $kn$ , the running time order of the greedy algorithm. It can be observed that EPORSS takes only about  $0.9kn$  and  $0.6kn$  iterations to achieve the best performance on the two data sets, respectively, implying that EPORSS can be more efficient in practice.

## 7 Conclusion

In this paper, we study the robust subset selection problem with monotone objective functions, and give two algorithms, i.e., the greedy algorithm and the evolutionary Pareto optimization algorithm EPORSS, with bounded approximation guarantees. The experimental results on the application of robust influence maximization show that the greedy algorithm can achieve competitive performance to the previous algorithms, but is more efficient; while EPORSS can achieve better performance by using more time. Thus, these two algorithms can help solve various real-world robust subset selection problems better. Note that we have considered the worst-case running time complexity of EPORSS in theoretical analysis, and the empirical results show that EPORSS can be efficient in practice. Thus, it would be interesting to analyze the average-case running time complexity, which may reflect the practical performance of EPORSS more accurately.

## References

- [Anari *et al.*, 2019] N. Anari, N. Haghtalab, J. Naor, S. Pokutta, M. Singh, and A. Torrico. Structured robust submodular maximization: Offline and online algorithms. In *AISTATS*, pages 3128–3137, Naha, Japan, 2019.
- [Bian *et al.*, 2017] A. A. Bian, J. M. Buhmann, A. Krause, and S. Tschachtschek. Guarantees for greedy maximization of non-submodular functions with applications. In *ICML*, pages 498–507, Sydney, Australia, 2017.
- [Chen *et al.*, 2009] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, pages 199–208, Paris, France, 2009.
- [Chen *et al.*, 2016] W. Chen, T. Lin, Z. Tan, M. Zhao, and X. Zhou. Robust influence maximization. In *KDD*, pages 795–804, San Francisco, CA, 2016.
- [Das and Kempe, 2011] A. Das and D. Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *ICML*, pages 1057–1064, Bellevue, WA, 2011.
- [Elenberg *et al.*, 2018] E. R. Elenberg, R. Khanna, A. G. Dimakis, and S. Negahban. Restricted strong convexity implies weak submodularity. *The Annals of Statistics*, 46(6B):3539–3568, 2018.
- [Feng *et al.*, 2019] C. Feng, C. Qian, and K. Tang. Unsupervised feature selection by Pareto optimization. In *AAAI*, pages 3534–3541, Honolulu, HI, 2019.
- [Friedrich and Neumann, 2015] T. Friedrich and F. Neumann. Maximizing submodular functions under matroid constraints by evolutionary algorithms. *Evolutionary Computation*, 23(4):543–558, 2015.
- [Friedrich *et al.*, 2019] T. Friedrich, A. Göbel, F. Neumann, F. Quinzan, and R. Rothenberger. Greedy maximization of functions with bounded curvature under partition matroid constraints. In *AAAI*, pages 2272–2279, Honolulu, HI, 2019.
- [Goyal *et al.*, 2011] A. Goyal, W. Lu, and L. Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *ICDM*, pages 211–220, Vancouver, Canada, 2011.
- [Harshaw *et al.*, 2019] C. Harshaw, M. Feldman, J. Ward, and A. Karbasi. Submodular maximization beyond non-negativity: Guarantees, fast algorithms, and applications. In *ICML*, pages 2634–2643, Long Beach, CA, 2019.
- [Hassidim and Singer, 2017] A. Hassidim and Y. Singer. Robust guarantees of stochastic greedy algorithms. In *ICML*, pages 1424–1432, Sydney, Australia, 2017.
- [He and Kempe, 2016] X. He and D. Kempe. Robust influence maximization. In *KDD*, pages 885–894, San Francisco, CA, 2016.
- [Hou and Clark, 2021] Q. Hou and A. Clark. Robust maximization of correlated submodular functions under cardinality and matroid constraints. *IEEE Transactions on Automatic Control*, 66(12):6148–6155, 2021.
- [Iyer, 2019] R. Iyer. A unified framework of robust submodular optimization. *arXiv:1906.06393*, 2019.
- [Kalimeris *et al.*, 2019] D. Kalimeris, G. Kaplun, and Y. Singer. Robust influence maximization for hyperparametric models. In *ICML*, pages 3192–3200, Long Beach, CA, 2019.
- [Kempe *et al.*, 2003] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, Washington, DC, 2003.
- [Krause *et al.*, 2008a] A. Krause, M. H. Brendan, G. Carlos, and A. Gupta. Robust submodular observation selection. *Journal of Machine Learning Research*, 9:2761–2801, 2008.
- [Krause *et al.*, 2008b] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.
- [Lin and Bilmes, 2011] H. Lin and J. Bilmes. A class of submodular functions for document summarization. In *ACL*, pages 510–520, Portland, OR, 2011.
- [Minoux, 1978] M. Minoux. Accelerated greedy algorithms for maximizing submodular set functions. *Optimization Techniques*, pages 234–243, 1978.
- [Mirzasoileiman *et al.*, 2015] B. Mirzasoileiman, A. Badanidiyuru, A. Karbasi, J. Vondrák, and A. Krause. Lazier than lazy greedy. In *AAAI*, pages 1812–1818, Austin, TX, 2015.
- [Nemhauser and Wolsey, 1978] G. L. Nemhauser and L. A. Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*, 3(3):177–188, 1978.
- [Nemhauser *et al.*, 1978] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions – I. *Mathematical Programming*, 14(1):265–294, 1978.
- [Qian *et al.*, 2015] C. Qian, Y. Yu, and Z.-H. Zhou. Subset selection by Pareto optimization. In *NIPS*, pages 1765–1773, Montreal, Canada, 2015.
- [Qian *et al.*, 2016] C. Qian, J.-C. Shi, Y. Yu, K. Tang, and Z.-H. Zhou. Parallel Pareto optimization for subset selection. In *IJCAI*, pages 1939–1945, New York, NY, 2016.
- [Qian *et al.*, 2018] C. Qian, Y. Yu, and K. Tang. Approximation guarantees of stochastic greedy algorithms for subset selection. In *IJCAI*, pages 1478–1484, Stockholm, Sweden, 2018.
- [Qian *et al.*, 2019] C. Qian, Y. Yu, K. Tang, X. Yao, and Z.-H. Zhou. Maximizing submodular or monotone approximately submodular functions by multi-objective evolutionary algorithms. *Artificial Intelligence*, 275:279–294, 2019.
- [Thompson, 2022] R. Thompson. Robust subset selection. *Computational Statistics and Data Analysis*, 169(107415):1–19, 2022.
- [Udwani, 2018] R. Udwani. Multi-objective maximization of monotone submodular functions with cardinality constraint. In *NIPS*, pages 9513–9524, Montréal, Canada, 2018.
- [Zhu *et al.*, 2015] F. Zhu, B. Fan, X. Zhu, Y. Wang, S. Xiang, and C. Pan. 10,000+ times accelerated robust subset selection (ARSS). In *AAAI*, pages 3217–3223, Austin, TX, 2015.