The Facility Location Problem with Aleatory Agents

Gennaro Auricchio
University of Padua
Padua, Italy
gennaro.auricchio@unipd.it

Jie Zhang
University of Bath
Bath, United Kingdom
jz2558@bath.ac.uk

ABSTRACT

In this paper, we introduce and study the Facility Location Problem with Aleatory Agents (FLPAA), where the facility can accommodate a number of agents, denoted as n, which is larger than the number of agents reporting their preferences, denoted as n_r. The spare capacity is used by $n_u := n - n_r$ aleatory agents distributed according to μ. The goal of FLPAA is to find a location y that minimises the *ex-ante social cost*, defined as the expected cost of the n_u agents sampled from μ plus the classic social cost incurred by the agents reporting their position. We investigate the mechanism design aspects of the FLPAA under the assumption that the mechanism designer lacks knowledge of the distribution μ but can query k quantiles of μ. We explore the trade-off between acquiring more insights into the probability distribution and designing a better-performing mechanism, which we describe through the strong approximation ratio (SAR). The SAR of a mechanism measures the highest ratio between the cost of the mechanism and the cost of the optimal solution on the worst-case input $\vec{x}$ and worst-case distribution μ, offering a stringent metric that does not depend on μ. In most cases, the lower bound matches the upper bound, proving that our mechanisms are tight, thus no truthful mechanism can achieve a lower SAR. Lastly, we extend our study to the case in which we must locate two facilities.

KEYWORDS

Algorithmic Game Theory; Mechanism Design; Facility Location Problem; Strong Approximation Ratio

ACM Reference Format:

Gennaro Auricchio and Jie Zhang. 2026. The Facility Location Problem with Aleatory Agents. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/HKJY8780

1 INTRODUCTION

The Facility Location Problem (FLP) is a classic problem in combinatorial optimisation whose objective is to determine the optimal placement of facilities to minimise transportation costs associated with servicing customers [27]. Ever since its introduction, the FLP has become a key subproblem in several social choice-related topics, such as disaster relief [9], supply chain management [31], healthcare [1], clustering [26], and public facilities accessibility [10]. Due to its practical significance, the facility location problems has garnered attention across diverse fields as operations research, theoretical computer science, economics, and computational game theory [14].

In economics and computational game theory, the study of the FLP has a distinct perspective. Instead of finding an algorithm to compute a solution, the interest is in defining a routine that elicits the position of a facility from the information reported by strategic *agents*. Since every agent needs to access a facility, they will misrepresent their information if this leads the routine to place the facilities in a preferred position, adding a layer of complexity to the FLP. Indeed, optimising an objective relying on the reported agents' preferences leads to undesirable manipulation fuelled by the agents' selfishness. Hence, a characteristic that mechanisms must possess is *truthfulness*, which ensures that no agent benefits by misrepresenting their private information. Committing to truthful routines, however, leads to suboptimal solutions. The standard value to quantify the trade-off between achieving the optimal objective and implementing a truthful mechanism is the *approximation ratio* – the worst-case ratio between the objective achieved by the mechanism and the optimal objective attainable [34]. Since a higher approximation ratio indicates a greater deviation from the optimal solution, the defining challenge in Algorithmic Mechanism Design is then to define truthful routines with a small approximation ratio.

Motivation of our model. In the classic FLP problem, the facility location is decided only based on the agents' reports. This assumption is limiting, as facilities such as hospitals or bus stops are open to everyone, not just the agents engaged in the eliciting procedure. Therefore, optimising solely on the reports of interested agents without considering external participants does not locate the facility at the best possible place. In fact, consider the case where a chain of coffee shops wants to open a new branch on a street. Some of the people living in the street are already customers of the said chain of coffee shops, thus they engage in the process of eliciting the position of the new branch by reporting their preferences on where the new shop should be placed. However, the number of agents that have already been customers of the shop is smaller than the total number of agents the shop can serve. In this case, it is necessary to keep in mind that new customers living in the street will use the facility after its opening. Denoting with μ the population density of the street, the mechanism designer has to define a routine that minimises the cost of the agents reporting their position while considering that new customers drawn from μ will access the facility as well. Similarly, those advocating for a new bus stop or a new hospital may reflect only a specific subset of the population–such as individuals already aware of the proposal–rather than the population as a whole.

As these examples showcase, eliciting the location of a facility based solely on agents' reports is limiting, as there is no guarantee that the set of reporting agents accurately represents the underlying population density μ. Furthermore, the mechanism designer may be

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/HKJY8780

unable to access complete information about the population density due to privacy constraints, limited data availability, or insufficient insight. This motivates the design of mechanisms that operate with minimal information about μ, such as knowledge of only the median or key quantiles of μ. To address these challenges, we introduce the Facility Location Problem with Aleatory Agents (FLPAA) and investigate its algorithmic and mechanism design aspects.

Our Contribution and Technique Overview. In this paper, we develop a novel framework for the Facility Location Problem (FLP) where the number of agents the facility can accommodate, namely n, is larger than the number of agents reporting their preferences, namely n_r. We assume that the spare capacity of the facility, namely $n_u := n - n_r$, is used by external agents modelled as independent and identically distributed (i.i.d.) samples of a probability distribution $\mu \in \mathcal{P}(\mathbb{R})$, where $\mathcal{P}(\mathbb{R})$ denotes the set of probability measures over $\mathbb{R}$. To evaluate the quality of a facility location y, we introduce the *ex-ante social cost*, which measures the expected cost of n_u i.i.d. agents distributed according to μ added to the social cost of the agents reporting their position

$$\mathcal{ESC}(\vec{x}, y; \mu) := \sum_{i=1}^{n_r} |x_i - y| + \sum_{j=1}^{n_u} \mathbb{E}_{X_j \sim \mu}[|X_j - y|]. \qquad (1)$$

Given $n \in \mathbb{N}$, $\vec{x} \in \mathbb{R}^{n_r}$ where $n_r \leq n$ and $\mu \in \mathcal{P}(\mathbb{R})$, the Facility Location Problem with Aleatory Agents (FLPAA) consists in finding $y \in \mathbb{R}$ that minimises the objective in (1).

First, we fully characterise the set of optimal solutions of the FLPAA for any given μ and $\vec{x}$. We show that to retrieve the optimal solution, it is not necessary to have access to a full description of μ, but rather to n_u carefully chosen quantiles of the distribution. We then study the mechanism design aspects of the FLPAA. Please notice that, while the first set of n_r agents report their preferences to a mechanism, the remaining n_u agents do not submit their preferences to the mechanism. Therefore, the mechanisms need to be truthful only with respect to the reports of the n_r agents, while the total cost has to be computed with respect to all the $n_r + n_u$ agents. We consider the case in which the mechanism designer does not know μ, but they can query any $k \in \mathbb{N}$ quantiles of μ. This allows us to outline the trade-off between gathering more insight on μ and defining a better-performing truthful mechanism. We introduce the notion of strong approximation ratio (SAR), which evaluates the mechanism on the worst-case input $\vec{x}$ and the worst-case distribution μ , i.e. $\mathsf{SAR}(M) := \sup_{\mu \in \mathcal{P}(\mathbb{R})} \sup_{\vec{x} \in \mathbb{R}^{n_r}} \frac{\mathcal{ESC}_M(\vec{x};\mu)}{\mathcal{ESC}_{opt}(\vec{x};\mu)}$, where $\mathcal{ESC}_M(\vec{x}; \mu) := \mathcal{ESC}(\vec{x}, M(\vec{x}); \mu)$ denotes the ex-ante social cost achieved by a truthful mechanism M and $\mathcal{ESC}_{opt}(\vec{x}; \mu) := \min_{y \in \mathbb{R}} \mathcal{ESC}(\vec{x}, y; \mu)$ is the optimal ex-ante social cost. Notice that the SAR is a stricter metric than the classic approximation ratio, as it provides efficiency guarantees that do not depend on μ. Throughout our study, we focus our attention on the set of Phantom Quantile Mechanisms (PQM), a family of truthful mechanisms that take as input the agents' reports and, depending on the quantiles available to the mechanism designer, return a location. We study the SAR guarantees of the PQM depending on the number of quantiles that the mechanism designer can query. We divide our investigation into four information settings determined by the value of k: (i) the *zero information case*, in which the mechanism designer cannot query any quantile of μ; (ii) the *median case*, in which the mechanism designer has access only to the median of μ; (iii) the *k-quantiles case*, in which the mechanism designer has access to k quantiles with $k = 2, \dots, n_u$; and (iv) the *n_u-quantiles information case*, where the mechanism designer can query for at least n_u quantiles of μ. For each information setting, we provide an upper and lower bound on the SAR attainable by truthful mechanisms and characterise the PQM attaining the minimal SAR. In Table 1, we summarise our results in terms of upper and lower bounds. Notice that the lower and upper bounds coincide when $k = 0$, $k = 1$ and the ratio between the number of agents reporting their positions is smaller than $1/3$ and when k is larger than n_u. Lastly, we extend our framework to the case in which we must locate two facilities capable of accommodating c agents. In each scenario, we define a truthful mechanism with bounded SAR or demonstrate that no truthful mechanism has bounded SAR. Due to space limits, the full version of the proofs is reported in the appendix.

Related Works. The study of the mechanism design aspects of the FLP was introduced in [34, 37]. Following that, a range of mechanisms with constant approximation ratios for placing one or two facilities on trees [22], circles [28], general graphs [2], and metric spaces [30] were introduced. Moreover, different works tried to generalise the problem by considering different agents' preferences [17, 23, 29], different types of facilities [20, 29], different costs [21], and additional constraints [42]. The m-Capacitated Facility Location Problem (m-CFLP) is a variant of the m-FLP where each facility can accommodate a finite number of agents [12, 35]. Considering facilities with capacity constraints is a natural approach for modelling scenarios where facilities offer limited resources, as it happens in distribution planning [36] and telecommunication network design [11, 15]. Within this framework, it was proven that no mechanism with a finite approximation ratio can place more than two capacitated facilities while adhering to truthfulness, anonymity, and Pareto optimality [39]. Recently, this impossibility result has been overcome by dropping Pareto optimality [4, 5]. Lastly, in [3, 8] the authors studied the m-CFLP under the assumption that the total capacity of the facility is not sufficient to accommodate all agents.

To some extent, our framework is similar to Bayesian Mechanism design, where agents' preferences are distributed according to μ. Under these assumptions, the designer has access to more information and thus it is able to refine its approximation guarantees. Bayesian Mechanism Design has been applied to investigate routing games [24], facility location problems [6, 41], and auction mechanism design [16, 19, 25, 40]. Our framework distinguishes itself from Bayesian mechanism design for two reasons: (i) In our case, the mechanism designer does not know the probability distribution μ, but only some qualitative information of μ, *e.g.* the quantiles. (ii) In Bayesian mechanism design, the performances of the mechanisms are measured as the ratio of the expectations [25] or as the expectation of the ratio [18]. In our case, the SAR is defined as a worst-case ratio. Finally, some work proposed models for the FLP which involve a degree of uncertainty. In [13], the authors explored the FLP when agents' locations are i.i.d. samples drawn from an unknown distribution. A different approach is presented in [32], where each agent is associated with an interval on the line that represents all its possible locations.

	Lower Bound	Upper Bound
$k=0$	$\frac{2}{\lambda}-1$	$\frac{2}{\lambda}-1$
$k=1$	$\begin{cases}\max\left\{\frac{4}{1+\lambda},2\right\}-1 & \text{if } \lambda\geq\frac{1}{3}\\ 1+\frac{2\lambda}{1-\lambda} & \text{otherwise}\end{cases}$	$\begin{cases}\max\left\{\frac{2}{\lambda+\frac{1}{n}},2\right\}-1 & \text{if } \lambda\geq\frac{1}{2}\\ 1+\frac{2\lambda}{1-\lambda} & \text{otherwise}\end{cases}$
$1<k<n_u$	$1+\dfrac{2(1-\lambda)\sigma}{(1+\lambda)n_u+(1-\lambda)\sigma}$	$1+\dfrac{2(1-\lambda)(\sigma-1)}{n_u-(1-\lambda)(\sigma-1)}$
$k\geq n_u$	1	1

Table 1: A table containing the upper and lower bounds on the Strong Approximation Ratio for all four frameworks of the one facility case we considered. The value n represents the number of agents the facility can serve, n_r the number of agents reporting their position, and $n_u := n - n_r$. The value $\lambda = \frac{n_r}{n}$ is the fraction of agents reporting their position. For the sake of simplicity, in the case $1 < k < n_u$, we assume that k divides n_u, that is $n_u = \sigma k$ where $\sigma \in \mathbb{N}$. Moreover, the lower bound we reported is restricted to the case in which k is even and the quantiles are induced by a vector $\vec{q} \in [0,1]^k$ whose entries are equi-distanced, that is $q_j = \frac{2j-1}{2k}$.

2 SETTING STATEMENT

In what follows, we assume that agents and the facility are placed on a line. Let n denote the capacity of the facility, n_r denote the number of agents reporting their position, and $n_u = n - n_r$ denote the spare capacity of the facility used by agents belonging to the population μ. We name *deterministic agents* the agents reporting their position, while the other agents are called *aleatory agents.*

Given the position of the n_r deterministic agents, our goal is to place a facility in a position that minimises the combined costs of the n_r deterministic agents and the expected costs of n_u i.i.d. aleatory agents distributed according to $\mu \in \mathcal{P}(\mathbb{R})$. In our framework, we assume that the designer does not have access to the whole distribution, but it can query a fixed number of quantiles. The quantity of quantiles to query is fix, but the mechanism designer decides which quantiles to query for. The quantiles that are queried are public information, thus the agents have access to them. Given a position $y \in \mathbb{R}$, a deterministic agent located at x_i incurs a cost of $c_i(x_i, y) = |x_i - y|$ to access the facility, while an aleatory agent sampled from $X \sim \mu$ incurs in an ex-ante cost equal to $c(y,\mu) = \mathbb{E}_{X\sim\mu}[|X - y|]$.

Problem 1. *Let $n = n_r + n_u$ be the capacity of a facility, $\mu \in \mathcal{P}(\mathbb{R})$, and $\vec{x} \in \mathbb{R}^{n_r}$. The point $y^* \in \mathbb{R}$ is a solution to the Facility Location Problem with Aleatory Agents (FLPAA) induced by $\vec{x}$ and μ if it minimises the ex-ante social cost function, namely $\mathcal{ESC}$, that is*

$$y \to \mathcal{ESC}(\vec{x}, y; \mu) = \sum_{i=1}^{n_r} |x_i - y| + n_u \mathbb{E}_{X\sim\mu}[|X - y|]. \tag{2}$$

Basic Assumptions. In what follows, we tacitly assume that the underlying distribution μ satisfies the two following properties: (i) $\mu \in \mathcal{P}(\mathbb{R})$ has finite first moment, i.e. $\int_{\mathbb{R}} |x| d\mu < +\infty$. This condition is essential, as otherwise the expected ex-ante cost of the aleatory agents is not finite. (ii) The measure μ is absolutely continuous. We denote with ρ_μ its probability density function, with F_μ its cumulative distribution function (c.d.f.), and with $F_\mu^{[-1]}$ its pseudo-inverse function. This assumption simplifies the analysis, and it is not stringent since any probability measure can be approximated by an absolutely continuous one [38], a result that we repeatedly use in our discussion. We say that a sequence of probability measures μ_ℓ *concentrates the probability at* one or more points $a \in \mathcal{A} \subset \mathbb{R}$ as $\ell \to \infty$ if μ_ℓ converges in distribution to a discrete probability measure supported over $\mathcal{A}$. An example is given by the sequence $\mu = 2\ell\mathcal{U}_{[x-\frac{1}{\ell}, x+\frac{1}{\ell}]}$, which converges to a probability measure that assigns probability 1 to $x \in \mathbb{R}$, thus the sequence μ_ℓ concentrates all the probability at x. In Figure 1, we give a graphical description of what concentrating the probability means.

2.1 The Optimal Solution of the FLPAA

In this section, we study the optimal solution to Problem 1. Given a vector $\vec{x} = (x_1, \ldots, x_{n_r})$, the *empirical cumulative distribution function* (ecdf) associated with $\vec{x}$ is defined as $F_{\vec{x}}(t) = \frac{1}{n_r}\sum_{i=1}^{n_r} \mathbb{I}_{\{x_i \leq t\}}(t)$, where $\mathbb{I}_{\{x_i\leq t\}}(t)$ is the indicator function of the set $\{x_i \leq t\}$, which is equal to 1 if $x_i \leq t$ and 0 otherwise. Given $\frac{n_r}{n} = \lambda \in [0,1]$, we set

$$F_{\lambda,\mu,\vec{x}}(t) = \lambda F_{\vec{x}}(t) + (1-\lambda)F_\mu(t), \tag{3}$$

where F_μ is the c.d.f. of μ. Since $F_{\lambda,\mu,\vec{x}}$ is the convex combination of two c.d.f., $F_{\lambda,\mu,\vec{x}}$ is a c.d.f. as well. We start our discussion by characterising the set containing the optimal solutions to the FLPAA.

Lemma 1. *Given $\vec{x} \in \mathbb{R}^{n_r}$ and $\mu \in \mathcal{P}(\mathbb{R})$, every median of $F_{\lambda,\mu,\vec{x}}$ is a solution, thus $y^* = \inf\left\{y \in \mathbb{R}, \text{ s.t. } F_{\lambda,\mu,\vec{x}}(t) \geq \frac{1}{2}\right\}$ is a solution. Moreover, given $\vec{x} \in \mathbb{R}^{n_r}$ and $\mu \in \mathcal{P}(\mathbb{R})$, $y \in \mathbb{R}$ is a solution to Problem 1 if and only if $y \in [a,b]$ where $a = \sup\{t \in \mathbb{R}, \text{ s.t. } \partial_y \mathcal{ESC}(t) < 0\}$ and $b = \inf\{t \in \mathbb{R}, \text{ s.t. } \partial_y \mathcal{ESC}(t) > 0\}$.*

From Lemma 1 we infer that the FLPAA has at least a solution. Moreover, there exists a discrete set that contains at least one solution to the FLPAA that is characterisable as follows.

Theorem 1. *Given n, n_r, n_u and μ, let us denote with $\mathcal{F}_{n_u} = \{f_j\}_{j\in[n_u]}$ where $f_j = F_\mu^{[-1]}(\frac{2j-1}{2n_u})$ if n is odd and $f_j = F_\mu^{[-1]}(\frac{j}{n_u})$ if n is even. Then, for any given $\vec{x} \in \mathbb{R}^{n_r}$, at least one element of the set $\mathcal{X} \cup \mathcal{F}_{n_u}$ is an optimal solution to Problem 1, where $\mathcal{X} = \{x_i\}_{i\in[n_r]}$. In particular, given $\mathcal{F}_{n_u}$, Problem 1 is solvable in polynomial time.*

Notice that the values defining $\mathcal{F}_{n_u}$ do not depend on μ. Indeed, $f_j \in \mathcal{F}_{n_u}$ is the quantile associated to the value $\frac{2j-1}{2n_u}$ regardless of μ. By refining the proof of Theorem 1, it is possible to restrict the set containing the possible optimal solutions whenever $n_r < n_u$.

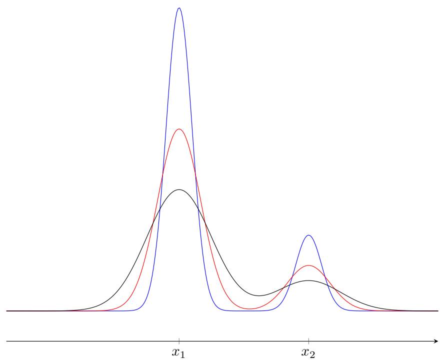

Figure 1: A sequence of probability distributions that concentrates the probability at x_1 and x_2. In the limit, the sequence converges to a probability measure supported on x_1 and x_2.

Corollary 1. *If $n_r < n_u$, the set $\mathcal{X}\cup\tilde{\mathcal{F}}_{n_u}$, where $\tilde{\mathcal{F}}_{n_u} = \{f_j\}_{j\in\{\alpha,\dots,\beta\}}$ with $\alpha = \frac{n+1}{2} - n_r$ and $\beta = n_u - \alpha$ contains at least an optimal solution.*

We define the set of relevant quantiles $\mathcal{R}(n_r, n_u)$ as the set of indexes $j \in [n_u]$ for which there exists $\vec{x} \in \mathbb{R}^{n_r}$ and $\mu \in \mathcal{P}(\mathbb{R})$ such that the optimal solution to Problem 1 is $F_\mu^{[-1]}(\frac{2j-1}{2n_u})$. More formally, when n is odd, we have

$$\mathcal{R}(n_r, n_u) := \Big\{ j \in [n_u] \ \text{ s.t. } \exists \vec{x} \in \mathbb{R}^{n_r}, \quad F_\mu^{[-1]}\Big(\frac{2j-1}{2n_u}\Big) \in \arg\min_{t\in\mathbb{R}} \Big\{ F_{\lambda,\mu,\vec{x}}(t) \geq \frac{1}{2} \Big\}\Big\}. \tag{4}$$

Similarly, we define $\mathcal{R}(n_r, n_u)$ for n even, in both cases, the cardinality of $\mathcal{R}(n_r, n_u)$ is finite. As we will see, the set $\mathcal{R}(n_r, n_u)$ contains all the information needed by the mechanism designer to define an optimal and truthful mechanism.

2.2 Mechanism Design Aspects of the FLPAA

A mechanism is a function $M : \mathbb{R}^{n_r} \to \mathbb{R}$ that takes in input the reports of the n_r deterministic agents $\vec{x} \in \mathbb{R}^{n_r}$ and outputs a facility position $y = M(\vec{x}) \in \mathbb{R}$. A mechanism M is said to be *truthful* (or *strategy-proof*) if the cost of a deterministic agent is minimised when it reports its true position. That is, $c_i(x_i, M(\vec{x})) \leq c_i(x_i, M(\vec{x}_{-i}, x_i'))$ for any misreport $x_i' \in \mathbb{R}$, where $\vec{x}_{-i}$ is the vector $\vec{x}$ without its i-th component. Again, we stress that the mechanism has to be truthful with respect to the set of n_r deterministic agents, as they are the ones reporting the information before the facility gets located. The remaining n_u agents do not submit their preferences to the mechanism and cannot manipulate the outcome of the process.

Given a truthful mechanism M and $\mu \in \mathcal{P}(\mathbb{R})$, the approximation ratio of M with respect to the $\mathcal{ESC}$ is $ar_\mu(M) := \sup_{\vec{x}\in\mathbb{R}^{n_r}} \frac{\mathcal{ESC}_M(\vec{x};\mu)}{\mathcal{ESC}_{opt}(\vec{x};\mu)}$. Notice that the approximation ratio of M depends on the probability measure μ. We are interested in defining routines whose performance is not dependent on μ, but that work well across all possible probability distributions. For this reason, we introduce the notion of strong approximation ratio (SAR), which measures the ratio of the cost attained by a mechanism and the optimal cost over every possible input $\vec{x}$ and over every possible distribution μ, that is

$$\mathsf{SAR}(M) := \sup_{\mu\in\mathcal{P}(\mathbb{R})} ar_\mu(M) = \sup_{\mu\in\mathcal{P}(\mathbb{R})} \sup_{\vec{x}\in\mathbb{R}^{n_r}} \frac{\mathcal{ESC}_M(\vec{x};\mu)}{\mathcal{ESC}_{opt}(\vec{x};\mu)}. \tag{5}$$

Notice that the SAR is a stricter metric than the classic approximation ratio since $ar_\mu(M) \leq \mathsf{SAR}(M)$. In our study, we assume that the mechanism designer does not know μ, but it can gather insight into μ by querying k quantiles of μ. These quantiles are set before the mechanism designer receives the agents' reports and thus are not manipulable by the agents. The challenge is then to define the best truthful mechanism M in terms of SAR based on the information on μ that the mechanism designer can gather. To this extent, we consider the class of *Phantom Quantile Mechanism* (PQM).

Definition 1. *Given $n_u \in \mathbb{N}$, let $\vec{q} \in [0,1]^{n_u}$ be such that $q_i \leq q_{i+1}$. Denoted with $F_\mu^{[-1]}$ the pseudo-inverse function of F_μ, the Phantom Quantile Mechanism (PQM) associated with $\vec{q}$ is defined as $\mathsf{PQM}_{\vec{q}}(\vec{x}) = med(\vec{x}, \vec{f})$, where $f_k = F_\mu^{[-1]}(q_k)$ for every $k \in [n_u]$.*

For any given $\vec{q}$, PQM is truthful and anonymous, as they are instances of the Phantom Peak Mechanisms, introduced in [33].

Theorem 2. *For every $\vec{q} \in [0,1]^{n_u}$ and every $\mu \in \mathcal{P}(\mathbb{R})$, the mechanism $\mathsf{PQM}_{\vec{q}}$ is truthful and anonymous.*

Notice that the PQM defines a class of mechanisms that depends on the quantiles $\vec{q}$. In what follows, we showcase how to select the best quantile vector $\vec{q}$ to query for any given number of agents reporting their own positions.

3 TAILOR-MAKING A MECHANISM TO DIFFERENT LEVELS OF INFORMATION

In what follows, we assume that the mechanism designer can query k quantiles from μ to tune the PQM. The k quantiles are queried before the mechanism designer receives the agents' reports, hence the agents cannot manipulate the way the designer selects the quantiles. We divide our study as follows: (i) in the *zero information case*, in which k is equal to 0; (ii) the *median information case*, in which the mechanism designer has access to the median of μ, *i.e.* $k = 1$, (iii) the *k-quantile information case*, in which the mechanism designer has access to any $1 < k < n_u$ quantiles of μ, and (iv) the *n_u-quantile information case*, in which $k \geq n_u$. Due to space limitations, we present only a sketch of the arguments for the most relevant theorems, to highlight the main techniques.

3.1 Zero Information case

In this case, the mechanism designer does not have access to any information on μ, thus the mechanism elicits the position of the facility based solely upon the agents' self-reported positions $\vec{x} \in \mathbb{R}^{n_r}$. In this case, placing the facility at the median of $\vec{x}$ defines a truthful mechanism that achieves bounded SAR.

Theorem 3. *The mechanism $M : \vec{x} \to med(\vec{x})$ is truthful. Moreover, we have that*

$$\mathsf{SAR}(M) = \begin{cases} \max\Big\{\frac{2n_u+n_r-1}{n_r+1}, 1\Big\} = \max\Big\{\frac{2-\lambda-\frac{1}{n}}{\lambda+\frac{1}{n}}, 1\Big\} & \text{if } n_r \text{ is odd,} \\ \frac{2n_u+n_r}{n_r} = \frac{2-\lambda}{\lambda} & \text{if } n_r \text{ is even.} \end{cases}$$

Sketch of the Proof. If $n_r = n$, the result follows trivially, we then consider the case $n_r < n$. Let $\vec{x} \in \mathbb{R}^{n_r}$ be the vector of agents' reports, increasingly ordered with distinct values. Denote by $y = M(\vec{x})$ the facility location selected by the median mechanism and by y^* the optimal location minimising (2). Without loss of generality, assume that $y^* = 0$ and $y < 0$.

Step 1: Upper bound. Using the triangle inequality, we bound the ex-ante social cost of the median mechanism as follows

$$\mathcal{ESC}_M(\vec{x},\mu) \leq \sum_{i=1}^{n_r} |x_i - y| + n_u\big(\mathbb{E}[|X|] + |y|\big).$$

Since $y^* = 0$, we have $\mathcal{ESC}_{opt}(\vec{x},\mu) = \sum_{i=1}^{n_r} |x_i| + n_u\,\mathbb{E}[|X|]$, so the approximation ratio satisfies

$$\frac{\mathcal{ESC}_M(\vec{x},\mu)}{\mathcal{ESC}_{opt}(\vec{x},\mu)} \leq \frac{\sum_{i=1}^{n_r} |x_i - y| + n_u|y|}{\sum_{i=1}^{n_r} |x_i|}.$$

Maximising this ratio by placing half the agents at y and the rest at 0 yields $\frac{\mathcal{ESC}_M(\vec{x},\mu)}{\mathcal{ESC}_{opt}(\vec{x},\mu)} \leq \frac{\lfloor \frac{n_r-1}{2} \rfloor + n_u}{\lfloor \frac{n_r+1}{2} \rfloor}$.

Step 2: Tightness. Define a sequence of probability measures μ_ℓ with densities $\ell\,\rho_\mu(\ell x)$ so that $\mathbb{E}_{\mu_\ell}[|X|] \to 0$ and $\mathbb{E}_{\mu_\ell}[|X-y|] \to |y|$ as $\ell \to \infty$. Consider the instance in which $\lfloor \frac{n_r+1}{2} \rfloor$ agents are located at y and the rest at 0, we then have $\lim_{\ell\to\infty} \frac{\mathcal{ESC}_M(\vec{x},\mu_\ell)}{\mathcal{ESC}_{opt}(\vec{x},\mu_\ell)} = \frac{\lfloor \frac{n_r-1}{2} \rfloor + n_u}{\lfloor \frac{n_r+1}{2} \rfloor}$. Hence, the bound is tight and the result follows. □

We complement Theorem 3 with an example showcasing how the approximation ratio of a mechanism M for a fixed measure μ is smaller than the SAR of the mechanism.

Example 1. *Let us fix $n = 5$, $n_r = 3$, and $n_u = 2$. Owing to Theorem 3, mechanism $M : \vec{x} \to med(\vec{x})$ has a SAR equal to $\frac{3}{2}$. Consider the case in which μ is the uniform distribution on the interval $[1,2]$, that is $\mu = \mathcal{U}([1,2])$. Following the proof of Theorem 3, we infer that the worst case instance is $\vec{x} = (0,0,1.25)$, since $F_\mu^{[-1]}(\frac{1}{4}) = 1.25$. By a simple computation, we have $\mathcal{ESC}_{opt}(\vec{x}) = \frac{15}{4}$, while $\mathcal{ESC}_M(\vec{x}) = \frac{17}{4}$, hence $ar_{\mathcal{U}([1,2])}(M) = \frac{17}{15} \sim 1.13 < \frac{3}{2}$.*

To conclude the discussion on the Zero Information case, we show that the bound in Theorem 3 is tight. In particular, the median mechanism is the best possible truthful mechanism for this framework hence no truthful mechanism can achieve a lower SAR.

Theorem 4. *No truthful mechanism attains an SAR lower than $\max\{\frac{2-\lambda-\frac{1}{n}}{\lambda+\frac{1}{n}}, 1\}$ if n_r is odd or $\frac{2-\lambda}{\lambda}$ if n_r is even. Hence the median mechanism is the best mechanism for the Zero Information case.*

Sketch of the Proof. Consider the instance $\vec{x}$ with $x_i = -1$ if $i \leq \frac{n_r+1}{2}$ and $x_i = 0$ otherwise. Let M be a truthful mechanism returning $y = M(\vec{x})$; without loss of generality, assume $y \in \{-1, 0\}$ since otherwise (i) if $y < -1$ or $y > 0$, the SAR attained on the instance would be larger than SAR attained in the case $y \in [-1,0]$, and (ii) if $-1 < y < 0$, we exploit the truthfulness of M to reduce the analysis to the cases where, up to a constant, $y = -1$ or $y = 0$.

Case $y = -1$. Let $\mu_\ell = \ell\mathcal{U}_{[-\frac{1}{2\ell},\frac{1}{2\ell}]}$ be a sequence of uniform distributions. The optimal facility location y_ℓ^* lies in the support of μ_ℓ and converges to $y^* = 0$ as $\ell \to \infty$. Moreover,

$$\mathbb{E}_{\mu_\ell}[|X - y^*|] \leq \mathbb{E}_{\mu_\ell}[|X|] + \mathbb{E}_{\mu_\ell}[|y^*|] \leq \tfrac{2}{\ell} \to 0.$$

Hence,

$$\lim_{\ell\to\infty} \frac{\mathcal{ESC}_M(\vec{x};\mu_\ell)}{\mathcal{ESC}_{opt}(\vec{x};\mu_\ell)} = \frac{\lfloor \frac{n_r-1}{2} \rfloor + n_u}{\lfloor \frac{n_r+1}{2} \rfloor}.$$

Case $y = 0$. Using the translated distributions $\mu_\ell = \ell\mathcal{U}_{[-1-\frac{1}{2\ell},\frac{1}{2\ell}-1]}$, the same argument yields an equal or larger ratio than in the previous case, which concludes the proof. □

3.2 Median Information Case

We now consider the case in which the mechanism designer has only access to the median of the measure μ, which we denote with m. In this case, we consider the PQM induced by the vector $\vec{q} = (0.5,\dots,0.5) \in [0,1]^{n_u}$, that is $\mathtt{PQM}_{\vec{q}}(\vec{x}) = med(\vec{x},\vec{m})$, where $\vec{m} = (m,\dots,m) \in \mathbb{R}^{n_u}$. This mechanism is truthful and has bounded SAR, however, its SAR is always larger than 1 when $1 < n_u < n-1$.

Theorem 5. *Let $\vec{q} = (0.5,\dots,0.5)$. Then, the mechanism $\mathtt{PQM}_{\vec{q}}$ is optimal if and only if $n_u \in \{0,1,n\}$. In all other cases, we have*

$$\mathtt{SAR}(\mathtt{PQM}_{\vec{q}}) = \begin{cases} \max\left\{\frac{2}{\lambda+\frac{1}{n}}, 2\right\} - 1 & \text{if } \lambda \geq \frac{1}{2} \\ 1 + \frac{2\lambda}{1-\lambda} & \text{otherwise.} \end{cases}$$

In particular, $\mathtt{SAR}(\mathtt{PQM}_{\vec{q}}) \leq 3$.

To conclude this section, we prove a lower bound on the SAR for any truthful mechanism that has access to the median of μ.

Theorem 6. *Let M be a truthful mechanism that has access to the median of μ, then*

$$\mathtt{SAR}(M) \geq \begin{cases} \max\left\{\frac{4}{1+\lambda}, 2\right\} - 1 & \text{if } \lambda \geq \frac{1}{3} \\ \frac{2\lambda}{1-\lambda} + 1 & \text{otherwise.} \end{cases}$$

In particular, if $\lambda \geq \frac{n-1}{n}$, the lower bound is 1 and is matched by $\mathtt{PQM}_{\vec{q}}$ where $\vec{q} = (0.5,\dots,0.5)$.

3.3 k-Quantile Information Case

We now consider the case in which the mechanism designer has access to $1 < k < n_u$ quantiles. In particular, given $\vec{q} \in [0,1]^k$, we characterise the PQM that achieves the lowest SAR while having access to the quantiles associated with $\vec{q}$. We then characterise the optimal $\vec{q} \in [0,1]^k$. Notice that the PQM are defined for $\vec{q} \in [0,1]^{n_u}$, whereas in this framework we are able to query only for k quantiles. To overcome this issue, we introduce the lift operator, which returns a n_u dimensional vector starting from a k dimensional vector.

Definition 2. *Given n_u, n_r, and k, we define the lift operator $L : [0,1]^k \to [0,1]^{n_u}$ as follows*

$$L : \vec{q} \to L(q) := (\underbrace{q_1,\dots,q_1}_{t_1\text{-times}}, \underbrace{q_2,\dots,q_2}_{t_2\text{-times}}, \dots, \underbrace{q_k,\dots,q_k}_{t_k\text{-times}}),$$

where t_j is the number of elements of $\mathcal{R}(n_r,n_u)$ whose closest entry of $\vec{q}$ is q_j (break ties arbitrarily).

The lift operator allows us to retrieve a n_u dimensional vector from a k dimensional one, thus, given $\vec{q}$, the mechanism $\mathtt{PQM}_{L(\vec{q})}$ is well defined. In the following Lemma, we compute the SAR of $\mathtt{PQM}_{L(\vec{q})}$ and show that it is the best PQM induced by a vector that has the same entries as $\vec{q}$.

Lemma 2. *Given $\vec{q}$, the mechanism $\mathsf{PQM}_{L(\vec{q})}$ is well-defined. Moreover, we have that*

$$\mathsf{SAR}(\mathsf{PQM}_{L(\vec{q})}) = 1 + \frac{4(1-\lambda)\Delta_{n_r,n_u}(L(\vec{q}))}{1-2(1-\lambda)\Delta_{n_r,n_u}(L(\vec{q}))}, \tag{6}$$

where $\Delta_{n_r,n_u} : [0,1]^{n_u} \to \mathbb{R}$ is defined as

$$\Delta_{n_r,n_u}(\vec{w}) = \max_{j\in\mathcal{R}(n_r,n_u)} \left|w_j - \frac{2j-1}{2n_u}\right|,$$

and $\mathcal{R}(n_r,n_u)$ is the set of relevant quantiles defined in (4). Finally, $L(\vec{q})$ induces the PQM with the lowest SAR amid the class of PQM induced by vectors whose entries are the same as the entries of $\vec{q}$, that is $\mathsf{SAR}(\mathsf{PQM}_{L(\vec{q})}) \le \mathsf{SAR}(\mathsf{PQM}_{\vec{w}})$, for every $\vec{w} \in [0,1]^{n_u}$ such that, for every $i \in [n_u]$ there exists a $j \in [k]$ such that $w_i = q_j$.

Sketch of the Proof. Given $j \in \mathcal{R}(n_r,n_u)$, let $\vec{x}$ and μ be such that the optimal solution is $y^* = F_\mu^{[-1]}(\frac{2j-1}{2n_u})$ and $\mathsf{PQM}_{L(\vec{q})}(\vec{x}) = y = F_\mu^{[-1]}(q_j)$. Without loss of generality, we assume that $y < y^*$ and set $\Delta_q = \left|q_j - \frac{2j-1}{2n_u}\right|$. We then apply a sequence of modifications to the instance in order to maximise the ratio between the mechanism cost and the optimal cost. In particular, (i) we move all the deterministic agents on the left of y to y. Since we are decreasing the optimal and mechanism cost by the same quantity the ratio increases. Similarly, we increase the ratio by considering a sequence of probability measures that concentrate the probability that μ assigns to the left of y on a small interval close to y. (ii) We repeat the process in (i) to the agents and probability on the right of y^*. (iii) We move all the agents whose position is between y and y^* to y^*. Finally, we concentrate all the probability that μ assigns to (y, y^*) to y^*. Notice that all these modifications do not affect the position of the optimal solution nor the output of $\mathsf{PQM}_{L(\vec{q})}$. Lastly, we compute the optimal and the mechanism cost. With a slight abuse of notation, we denote with $\vec{x}$ and μ the agents' reports and the probability measure we obtained after modifying the instance according to points (i), (ii), and (iii). By Lemma 1, we have that $F_{\lambda,\mu,\vec{x}}(y^*) \ge \frac{1}{2}$. By construction, $F_{\lambda,\mu,\vec{x}}(y) = \frac{1}{2} - (1-\lambda)\Delta_q$, therefore the ratio between the mechanism cost and the optimal cost is $\frac{1+2(1-\lambda)\Delta_q}{1-2(1-\lambda)\Delta_q} = 1 + \frac{4(1-\lambda)\Delta_q}{1-2(1-\lambda)\Delta_q}$, which concludes the proof. ◻

We stress that Lemma 2 gives us a formula to retrieve the best PQM for any given quantile vector $\vec{q} \in [0,1]^k$. We leverage this result to characterise the vector $\vec{q}$ that minimises the SAR of $\mathsf{PQM}_{L(\vec{q})}$ for any given k, n_u, and n_r.

Theorem 7. *Given k, n_u, and n_r, let $\sigma, \tau \in \mathbb{N}$ be the unique pair of natural values such that $n_u = \sigma k + \tau$. If $\mathcal{R}(n_r,n_u) = \{\frac{2j-1}{2n_u}\}_{j\in[n_u]}$, the best PQM mechanism is $\mathsf{PQM}_{L(\vec{q})}$ where $\vec{q}$*

$$q_s = \frac{2(s-1)(\sigma+1)+\sigma+1}{2n_u} \qquad \text{if } s \le \sigma,$$

$$q_s = \frac{2(\tau-1)(\sigma+1)+2(s-\tau-1)\sigma+\sigma}{2n_u} \qquad \text{if } s > \sigma.$$

Owing to Theorem 7, when k divides n_u and $n_r > n_u$, the best PQM is induced by the vector $\vec{q} = (\frac{1}{2k}, \frac{3}{2k}, \ldots, \frac{2k-1}{2k})$ and its SAR is $1 + \frac{2(1-\sigma)(\tau-1)}{n_u-(1-\sigma)(\tau-1)}$, where $\sigma \in \mathbb{N}$ is such that $n_u = \sigma k$. Lastly, we provide a lower bound on the SAR of any truthful mechanism that places the facility while having access to the k quantiles induced by $\vec{q} = (\frac{1}{2k}; \frac{3}{2k}, \ldots, \frac{2k-1}{2k})$.

Theorem 8. *Given $k, \sigma \in \mathbb{N}$ such that $n_u = \sigma k$, let $\vec{q} \in [0,1]^k$ be the vector containing k equi-distanced values, that is $q_j = \frac{2j-1}{2k}$. Any truthful mechanism M that has access only to the quantiles induced by $\vec{q}$ is such that*

$$\mathsf{SAR}(M) \ge \begin{cases} 1 + \dfrac{2n_u}{k(n+n_u)-2n_u} & \text{if } k \text{ is even,} \\ 1 + \dfrac{6n_u}{k(n+n_u)-5n_u} & \text{otherwise.} \end{cases}$$

Sketch of the Proof. Let M be a truthful mechanism. For the sake of argument, let us assume that k and n_r are even. By definition of $\vec{q}$, we have that $q_{\frac{k}{2}} = \frac{1}{2} - \frac{1}{2k}$ and $q_{\frac{k}{2}+1} = \frac{1}{2} + \frac{1}{2k}$. Let μ be a probability measure such that $F_\mu^{[-1]}(q_{\frac{k}{2}}) = 0$, $F_\mu^{[-1]}(q_{\frac{k}{2}+1}) = 1$, $F_\mu^{[-1]}(q_j) \in (-\epsilon, 0)$ if $j < \frac{k}{2}$, and $F_\mu^{[-1]}(q_j) \in (1, 1+\epsilon)$ if $j > \frac{k}{2}+1$, where ϵ is a small positive constant. Let $\vec{x} \in \mathbb{R}^{n_r}$ be a vector such that $x_i = 0$ if $i \le \frac{n_r}{2}$ and $x_i = 1$ otherwise. Finally, let y denote the output of M for this instance, which we can assume to be in $[0,1]$, as otherwise the SAR of the mechanism is higher. Notice that we cannot restrict y to be either 0 or 1, as the truthfulness of M applies only to the agents' reports and not to the quantiles of μ. Indeed, to maximise the ratio between the mechanism cost and the optimal cost, we either (i) move all the agents located at 0 to y and concentrate all the probability that μ assigns to $(0,1)$ at 1; or (ii) move all the agents located at 1 to y and concentrate all the probability that μ assigns to $(0,1)$ at 0. Whether we modify the instance following (i) or (ii), depends on which modification leads to the highest ratio. The lower bound is then retrieved by selecting $y \in [0,1]$ that minimises the maximum ratio attainable by applying (i) or (ii). Owing to the symmetry of the instance, we infer that this happens when $y = \frac{1}{2}$. The full computation of the lower bounds is deferred to the Appendix. ◻

3.4 n_u-Quantile Information Case

When $k \ge n_u$, it is possible to define an optimal and truthful mechanism. Moreover, the routine of the optimal mechanism does not depend on the value of n_r. Indeed, given n_u, the PQM induced by the vector $\vec{q} = (\frac{1}{2n_u}, \frac{3}{2n_u}, \ldots, \frac{2n_u-1}{2n_u})$, is truthful and optimal.

Theorem 9. *The mechanism $\mathsf{PQM}_{\vec{q}}$ where $q_k = \frac{2k-1}{2n_u}$ for every $k \in [n_u]$ is truthful and optimal regardless of μ, that is $\mathsf{SAR}(M) = 1$.*

In particular, if the mechanism designer has access to the full distribution μ, they can design a truthful and optimal mechanism.

4 EXTENSION TO TWO FACILITIES

We extend our analysis to the setting with two capacitated facilities. Unlike the single-facility case, this scenario introduces additional complexity due to the fact that the mechanism must assign the agents to the facilities in such a way that the capacity limit is not violated. Moreover, impossibility results and optimality criteria for one facility do not directly generalise. In this section, we formally state the problem, address the issues arising in this setting, and derive mechanisms with bounded SAR when possible.

4.1 Setting Statement

We now consider the case where we have two facilities with capacity c to place, thus the total number of agents that the facilities can serve is $n = 2c$. We denote by $n_r \le n = 2c$ the number of agents reporting their position to the mechanism. Since we have two facilities with a capacity limit, the mechanism must elicit the positions of the facilities and then coordinate the agents by determining an agent-to-facility matching [7]. This ensures that no facility is overloaded.

Given the agent report $\vec{x}$, let y_1 and y_2 be the positions of the facilities and γ be the agent-to-facility matching. We denote with $n_u^{(1)}$ and $n_u^{(2)}$ the spare capacity of the facility at y_1 and the spare capacity of the facility at y_2, that is $n_u^{(j)} = c - \#\{i \in [n] \;\text{ such that }\; (i,j) \in \gamma\}$. To coordinate the aleatory agents, we need to determine a function $f_\gamma : \mathbb{R} \to \{y_1, y_2\}$ that maps the realisation of $X \sim \mu$ into the two facilities. Since y_1 has a spare capacity of $n_u^{(1)}$ and the facility at y_2 has a spare capacity of $n_u^{(2)}$, we define

$$f_\gamma : x \to \begin{cases} y_1 & \text{if} \quad x \le F_\mu^{[-1]}(\frac{n_u^{(1)}}{n_u}) \\ y_2 & \text{otherwise} \end{cases}. \tag{7}$$

Notice that f_γ is well and uniquely defined by $\vec{x}$, $\vec{y} = (y_1, y_2)$, γ, and μ. Finally, given the agents' reports $\vec{x}$ we define the expected ex-ante social cost associated with the tuple $(\vec{y}, \gamma, \mu)$ as

$$\mathcal{ESC}(\vec{x}; \vec{y}, \gamma, \mu) = \sum_{(i,j)\in\gamma} |x_i - y_j|\gamma_{i,j} + n_u \mathbb{E}_{X\sim\mu}[|X - f_\gamma(X)|]. \tag{8}$$

Given a vector $\vec{x}$ containing the deterministic agents' positions, the optimal 2-FLPAA is the tuple (y_1, y_2, γ) that minimises $\mathcal{ESC}$. We then denote with $\mathcal{ESC}_{opt}(\vec{x}; \mu)$ the minimum ESC attainable on instance $\vec{x}$, that is $\mathcal{ESC}_{opt}(\vec{x}; \mu) = \min_{\vec{y},\gamma} \mathcal{ESC}(\vec{x}; \vec{y}, \gamma, \mu)$. Similarly, we denote with $\mathcal{ESC}_M(\vec{x}; \mu)$ the ESC attained by the mechanism M, that is $\mathcal{ESC}_M(\vec{x}; \mu) = \mathcal{ESC}(\vec{x}; M(\vec{x}), \mu)$, where $M(\vec{x}) = (\vec{y}, \gamma)$.

Theorem 10. *Given $\vec{y} = (y_1, y_2)$ and γ, the function f_γ, defined as in* (7)*, has the following properties:*

(i) If we take n_u samples of X, the expected number of agents that f_γ assigns to y_1 is $n_u^{(1)}$. Similarly, the expected number of agents that f_γ assigns to y_2 is $n_u^{(2)}$, that is $n_u \mathbb{E}_{X\sim\mu}[\mathbf{Id}_{f(X)=y_j}] = n_u^{(j)}$ for $j = 1, 2$. Therefore the expected number of agents assigned to each facility is equal to the capacity of the facility.

(ii) Among the functions g satisfying property (i), f_γ minimises the expected ex-ante social cost of the allocation, that is $\mathbb{E}[|X - f(X)|] \le \mathbb{E}[|X - g(X)|]$ for every g satisfying property (i).

(iii) The function f_γ allows us to split μ into two measures with disjoint support, that is $\mu = F_\mu(z)\mu_{\le z} + (1 - F_\mu(z))\mu_{>z}$, where $z = F_\mu^{[-1]}(\frac{n_u^{(1)}}{n_u})$. Then, the expected ex-ante social cost is

$$\mathcal{ESC}(\vec{x}; \vec{y}, \gamma, \mu) = \sum_{(i,j)\in\gamma} |x_i - y_j| + n_u^{(1)} \mathbb{E}_{X_1\sim\mu_{\le z}}[|X_1 - y_1|] + n_u^{(2)} \mathbb{E}_{X_2\sim\mu_{>z}}[|X_2 - y_2|].$$

Given a facility location $\vec{y}$ and an agent-to-facility matching γ, the cost of a reporting agent is $c_i(x_i, \vec{y}) = |x_i - y_j|$, where (i,j) is the unique edge in γ adjacent to agent i. The ex-ante cost of an aleatory agent is $c_j(y_1, y_2) = \mathbb{E}_{X\sim\mu}[|X - f_\gamma(X)|]$, where f_γ is defined as in (7). As per the one facility case, the optimal solution to the 2-FLPAA has a closed form.

Theorem 11. *Given two facilities with capacity c and the agents' reports $\vec{x}$, let $F_{\lambda,\mu,\vec{x}}$ be defined as in* (3)*. If there exists $z \in \mathbb{R}$ such that $z = F_{\lambda,\mu,\vec{x}}^{[-1]}(0.5)$, the optimal solution to the 2-FLPAA is the tuple (y_1, y_2, γ) where (i) $y_1 = F_{\lambda,\mu,\vec{x}}^{[-1]}(0.25)$, $y_2 = F_{\lambda,\mu,\vec{x}}^{[-1]}(0.75)$; and (ii) $\gamma = \{(i,j)\}$ is defined as $(i,j) \in \gamma$ if and only if $x_i \le F_{\lambda,\mu,\vec{x}}^{[-1]}(0.5)$ and $j = 1$ or $x_i > F_{\lambda,\mu,\vec{x}}^{[-1]}(0.5)$ and $j = 2$. If there is no point $z \in \mathbb{R}$ such that $F_{\lambda,\mu,\vec{x}}(z) = 0.5$, γ assigns every agent whose position is strictly to the left of $z = F_{\lambda,\mu,\vec{x}}^{[-1]}(0.5)$ to y_1 and the agents whose position is strictly to the right of z to y_2. The agents at z are assigned to either y_1 or y_2 depending on which allocation induces the lowest cost.*

	$n_r \le c$	$n_r > c$
$k = 0$	N/A	$3(c-1)$
$1 \le k < n_u$	$3(c-1)$	$3(c-1)$
$k \ge n_u$	3	$3(c-1)$

Table 2: SAR guarantees attainable for the 2-FLPAA depending on the number of deterministic agents of the problem. In the Table, N/A stands for Not Applicable, meaning that no truthful mechanism attains bounded SAR. The value c represents the number of agents that each facility can serve, n_r the number of agents reporting their position, and $n_u = n - n_r$. For the sake of simplicity, in the case $1 < k < n_u$, we assume that k divides n_u, that is $n_u = \sigma k$ where $\sigma \in \mathbb{N}$. Moreover, the lower bound we reported is restricted to the case in which k is even and the quantiles are induced by a vector $\vec{q} \in [0,1]^k$ whose entries are equi-distanced, that is $q_j = \frac{2j-1}{2k}$.

As for the case of one facility, given μ, n_r, and c, the optimal position for the facilities belongs to a discrete set.

Corollary 2. *The optimal locations y_1 and y_2 belong to the following discrete set $\{x_i\}_{i\in[n_r]} \cup \{F_\mu^{[-1]}(\frac{2j-1}{2n_u})\}_{j\in[n_u]}$.*

4.2 Mechanism Design for the 2-FLPAA

We now study the 2-FLPAA from a mechanism design perspective. As per the one facility case, we divide the presentation into sections depending on how many quantiles the mechanism designer can query. Due to the nature of the problem, depending on the number and values that the mechanism designer queries, it might be impossible to define a truthful mechanism that attains bounded SAR. When this happens, we prove that there are no truthful mechanisms with bounded SAR. We summarise our findings in Table 2.

4.2.1 Zero Information Case. When the mechanism designer cannot query any quantile, it might be impossible to define a truthful mechanism with bounded SAR. Indeed, if $n_r \le c$, no truthful mechanism has bounded SAR.

Theorem 12. *Let M be a truthful mechanism, if $n_r \le c$, then* $\mathsf{SAR}(M) = +\infty$.

Proof. Let M be a truthful mechanism and let us consider the instance $\vec{x} = (1, 1, \dots, 1) \in \mathbb{R}^{n_r}$. Let us denote with $\vec{y} = (y_1, y_2) = M(\vec{x})$. Without loss of generality we can assume that $y_1, y_2 \ne 0$.

Let μ be the uniform distribution over the interval $[-1, 1]$, so that $\rho_\mu(t) = \frac{1}{2}$ if $t \in [-1, 1]$ and $\rho_\mu(t) = 0$ otherwise. We define

α_ℓ as the probability distribution whose density is $\ell\rho_\mu(\ell x)$ and β_ℓ as the density whose probability distribution is $\ell\rho_\mu(\ell(x-1))$. We then define $\mu_\ell = \frac{k}{n_u}\alpha_\ell + \frac{n_u-k}{n_u}\beta_\ell$. By the same argument used to prove Theorem 3, we have $\lim_{\ell\to\infty} \mathcal{ESC}_{opt}(\vec{x};\mu_\ell) = 0$, however $\lim_{\ell\to\infty} \mathcal{ESC}_M(\vec{x};\mu_\ell) \geq |y_1|$, which concludes the proof. □

When $n_r > c$, the InnerGap Mechanism, introduced in [39], is truthful, does not overload any facility, and attains bounded SAR.

Mechanism 1. *Given $\vec{x} \in \mathbb{R}^{n_r}$ with $n_r > c$, the InnerGap Mechanism (IGM) returns $y_1 = x_{n-c}$, $y_2 = x_{c+1}$, and assigns every agents to the facility that is closer to the position they reported.*

Theorem 13. *The IGM is truthful and* $\mathsf{SAR}(IGM) \leq 3(c-1)$.

4.2.2 n_u-quantiles Case. We now consider the case in which the mechanism designer has access to n_u quantiles of μ. Owing to Corollary 2, we focus our attention to the case in which the quantiles to query are the ones induced by $\vec{q} = (\frac{1}{2n_u}, \frac{3}{2n_u}, \dots, \frac{2n_u-1}{2n_u})$. Unfortunately, in this case, the optimal mechanism is not truthful.

Example 2. *Let us consider $c = 5$, so that the total capacity of the facilities is 10, let us set $n_r = 8$. Let μ be the uniform distribution on the interval $[0,1]$. Let $\vec{x} = (0,1,1,2,9,9,9,9) \in \mathbb{R}^8$ be the vector containing the agents' position. The optimal solution places y_1 at 0.75, y_2 at 9 and assigns the agents located at 0 and 1 to y_1, and all the others at y_2. The agent at 2 can manipulate. Indeed, if it manipulates by reporting 0.75 rather than 2, the mechanism still places y_1 at 0.75 and y_2 at 9, however, in this case the manipulative agent is assigned to y_1, thus its cost decreases from 7 to 1.25.*

Mechanism 2. *For every $c \in \mathbb{N}$ and $n_r \leq c$, let*

$$\vec{f} = \left(F_\mu^{[-1]}\Big(\frac{1}{2n_u}\Big), \dots, F_\mu^{[-1]}\Big(\frac{2n_u-1}{2n_u}\Big)\right).$$

Given $\vec{x}$ in input, the Pseudo Optimal Mechanism (POM) places the facilities at $z_{\lfloor \frac{c+1}{2} \rfloor}$ and $z_{n-\lfloor \frac{c}{2} \rfloor}$, where $\vec{z}$ is the vector obtained by ordering in increasing order the vector $(\vec{x}, \vec{f})$. Then POM assigns every agent to the facility that is closer to the position they report.

Theorem 14. *The POM is truthful and* $\mathsf{SAR}(POM) = 3$.

Notice that when $n_r > c$, the POM is no longer well-defined; we thus introduce a different mechanism to handle this case.

Mechanism 3. *Let $n_r > c$ and let $\vec{f}$ be the quantiles associated with $\vec{q} = (\frac{1}{2n_u}, \dots, \frac{2n_u-1}{2n_u})$. Given $\vec{x}$, let $\vec{z}$ be the vector obtained by reordering the entries of $(\vec{x}, \vec{f})$ increasingly. Then, the output of the Amended Quartiles Mechanism (AQM) on $\vec{x}$ is defined as follows (i) we define $\vec{y} = (y_1, y_2)$ where $y_1 = \max\{x_{n_r-c}, z_{\lceil \frac{c}{2} \rceil}\}$ and $y_2 = \min\{x_{c+1}, z_{n-\lfloor \frac{c}{2} \rfloor}\}$; and (ii) γ as $(i,j) \in \gamma$ if and only if $|x_i - y_j| = \min\{|x_i - y_1|, |x_i - y_2|\}$; that is every agent is assigned to the facility that is closer to their report.*

It is easy to see that the AQM does not overload any facility. Moreover, it is truthful and attains bounded SAR.

Theorem 15. *The AQM is truthful and* $\mathsf{SAR}(AQM) \leq 3(c-1)$.

4.2.3 k-quantile Case. To conclude, we study the case in which we have access to k quantiles of μ, where $k \in \{1, 2, \dots, n_u - 1\}$. Notice that this framework includes the one quantile case. First, we notice that, depending on the quantiles queried, there might be no truthful mechanism that has bounded SAR.

Example 3. *Let us fix $c = 3$, so that $n = 6$. We fix $n_r = 2$, thus $n_u = 4$. Assume that we query for 3 quantiles. If we query the quantiles associated with $q_1 = 0.05$, $q_2 = 0.1$, and $q_3 = 0.15$, any truthful mechanism has unbounded SAR. Indeed, let us consider the following instances indexed by $\ell \in \mathbb{N}$. The report of the agents are $x_1 = x_2 = 0$ for every ℓ, while $\mu_\ell = \frac{\ell}{4}\mathcal{U}_{(-\frac{1}{\ell},0)} + \frac{3\ell}{4}\mathcal{U}_{(T-\frac{1}{\ell},T)}$, where T is a parameter to fix. Given a truthful mechanism M, let $\vec{y}_\ell$ be the position at which it places the facilities on the instance. If we set $T \neq y_1, y_2$, we have that the optimal cost of the instance converges to zero, while the cost of the mechanism is always larger than 0.*

Despite the negative result in Example 3, querying the right quantiles overcomes this impossibility result: for $\vec{q} = \left(\frac{1}{2k}, \frac{3}{2k}, \dots, \frac{2k-1}{2k}\right)$, it is possible to design a truthful routine with bounded SAR.

Mechanism 4. *Let $\vec{q} = (\frac{1}{2k}, \dots, \frac{2k-1}{2k})$ and $\vec{f}$ be such that $f_k = F_\mu^{[-1]}(L(\vec{q})_k)$. For any $\vec{x} \in \mathbb{R}^{n_r}$, let $\vec{z} \in \mathbb{R}^n$ be the vector obtained by reordering $(\vec{x}, \vec{f})$. The Capped EndPoint Mechanism (CEM) is defined as follows. If $n_r \leq c$, then $CEM(\vec{x}) = (z_1, z_n)$. If $n_r > c$, $CEM(\vec{x}) = (\max\{x_{n-c}, z_{\lceil \frac{c}{2} \rceil}\}, \min\{x_{c+1}, z_{n-\lfloor \frac{c}{2} \rfloor}\})$. Every agent is then assigned to the facility closer to the position they reported.*

Since there are at most c agents on the right of y_1 and c agents on the left of y_2, CEM is well-defined and it does not overload any facility, moreover it is truthful and attains bounded SAR.

Theorem 16. *The CEM is truthful and* $\mathsf{SAR}(CEM) \leq 3(c-1)$.

5 CONCLUSION AND FUTURE WORK

In this paper, we introduced the Facility Location Problem with Aleatory Agents (FLPAA), where the facility accommodates agents whose position is known, along with agents whose position is aleatory. After characterising the optimal solution to the FLPAA for any given instance $\vec{x}$ and distribution μ, we studied the mechanism design aspects of the FLPAA. We considered the problem of designing truthful mechanisms that perform well when the mechanism designer does not have access to μ, but to k quantiles that the mechanism designer can query for. We introduced the notion of strong approximation ratio (SAR), which measures the ratio between the mechanism cost and the optimal cost on the worst-case input $\vec{x}$ and distribution μ. We studied the upper and lower bounds for every value of k and provided truthful routines with bounded SAR. In several cases, the upper bound matches the lower bound. Lastly, we extended our study to the case where we must locate two facilities with capacity c. For future works, we aim to improve the lower bound for the case in which $1 < k < n_u$, to extend our study of the FLPAA to higher dimensional spaces, and to a more generic class of costs, such as the Maximum or ℓ_p Costs [21].

ACKNOWLEDGMENTS

Jie Zhang was supported by a Leverhulme Trust Research Project Grant and the EPSRC grant (EP/W014912/1).

REFERENCES

[1] Amir Ahmadi-Javid, Pardis Seyedi, and Siddhartha S Syam. 2017. A survey of healthcare facility location. *Computers & Operations Research* 79 (2017), 223–263.

[2] Noga Alon, Michal Feldman, Ariel D. Procaccia, and Moshe Tennenholtz. 2010. Strategyproof Approximation of the Minimax on Networks. *Mathematics of Operations Research* 35, 3 (2010), 513–526.

[3] Gennaro Auricchio, Harry J Clough, and Jie Zhang. 2024. On the Capacitated Facility Location Problem with Scarce Resources. In *The 40th Conference on Uncertainty in Artificial Intelligence*.

[4] Gennaro Auricchio, Zihe Wang, and Jie Zhang. 2024. Facility Location Problems with Capacity Constraints: Two Facilities and Beyond. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*. 2651–2659.

[5] Gennaro Auricchio, Zihe Wang, and Jie Zhang. 2025. On the design of truthful mechanisms for the capacitated facility location problem with two and more facilities. *Artificial Intelligence* 348 (2025), 104390.

[6] Gennaro Auricchio, Jie Zhang, and Mengxiao Zhang. 2024. Extended Ranking Mechanisms for the m-Capacitated Facility Location Problem in Bayesian Mechanism Design. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*. 87–95.

[7] Haris Aziz, Hau Chan, Barton Lee, Bo Li, and Toby Walsh. 2020. Facility location problem with capacity constraints: Algorithmic and mechanism design perspectives. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. MIT Press, New York, 1806–1813.

[8] Haris Aziz, Hau Chan, Barton E Lee, and David C Parkes. 2020. The capacity constrained facility location problem. *Games and Economic Behavior* 124 (2020), 478–490.

[9] B. Balcik and B. M. Beamon. 2008. Facility location in humanitarian relief. *International Journal of Logistics Research and Applications* 11, 2 (2008), 101–121.

[10] O Haluk Barda, Joseph Dupuis, and Pierre Lencioni. 1990. Multicriteria location of thermal power plants. *European Journal of Operational Research* 45, 2-3 (1990), 332–346.

[11] TB Boffey. 1989. Location problems arising in computer networks. *Journal of the Operational Research Society* 40 (1989), 347–354.

[12] Jack Brimberg, Ephraim Korach, Moshe Eben-Chaim, and Abraham Mehrez. 2001. The Capacitated p-facility Location Problem on the Real Line. *International Transactions in Operational Research* 8, 6 (2001), 727–738.

[13] Ioannis Caragiannis, Ariel Procaccia, and Nisarg Shah. 2016. Truthful univariate estimators. In *International Conference on Machine Learning*. PMLR, 127–135.

[14] Derya Celik Turkoglu and Mujde Erol Genevois. 2020. A comparative survey of service facility location problems. *Annals of Operations Research* 292 (2020), 399–468.

[15] Pierre Chardaire. 1999. Hierarchical two level location problems. *Telecommunications network planning* (1999), 33–54.

[16] Shuchi Chawla, Jason D Hartline, and Robert Kleinberg. 2007. Algorithmic pricing via virtual valuations. In *Proceedings of the 8th ACM Conference on Electronic Commerce*. ACM, New York, 243–251.

[17] Richard L Church and Robert S Garfinkel. 1978. Locating an obnoxious facility on a network. *Transportation science* 12, 2 (1978), 107–118.

[18] Xiaotie Deng, Yansong Gao, and Jie Zhang. 2017. Smoothed and average-case approximation ratios of mechanisms: Beyond the worst-case analysis. *arXiv preprint arXiv:1705.07200* (2017).

[19] Ludwig Ensthaler and Thomas Giebe. 2014. Bayesian optimal knapsack procurement. *European Journal of Operational Research* 234, 3 (2014), 774–779.

[20] Itai Feigenbaum and Jay Sethuraman. 2015. Strategyproof mechanisms for one-dimensional hybrid and obnoxious facility location models. In *AAAI workshop: Incentive and trust in E-communities*.

[21] Itai Feigenbaum, Jay Sethuraman, and Chun Ye. 2017. Approximately optimal mechanisms for strategyproof facility location: Minimizing lp norm of costs. *Mathematics of Operations Research* 42, 2 (2017), 434–447.

[22] Michal Feldman and Yoav Wilf. 2013. Strategyproof facility location and the least squares objective. In *EC*. ACM, 873–890.

[23] Aris Filos-Ratsikas, Minming Li, Jie Zhang, and Qiang Zhang. 2017. Facility location with double-peaked preferences. *Auton. Agents Multi Agent Syst.* 31, 6 (2017), 1209–1235. https://doi.org/10.1007/S10458-017-9361-0

[24] Martin Gairing, Burkhard Monien, and Karsten Tiemann. 2005. Selfish routing with incomplete information. In *Proceedings of the seventeenth annual ACM symposium on Parallelism in algorithms and architectures*. ACM, New York, 203–212.

[25] Jason D Hartline and Tim Roughgarden. 2009. Simple versus optimal mechanisms. In *Proceedings of the 10th ACM conference on Electronic commerce*. ACM, 225–234.

[26] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, Berlin.

[27] Dorit S Hochbaum. 1982. Heuristics for the fixed cost median problem. *Mathematical programming* 22 (1982), 148–162.

[28] Pinyan Lu, Xiaorui Sun, Yajun Wang, and Zeyuan Allen Zhu. 2010. Asymptotically optimal strategy-proof mechanisms for two-facility games. In *EC*. ACM, New York, 315–324.

[29] Lili Mei, Deshi Ye, and Guochuan Zhang. 2016. Mechanism Design for One-Facility Location Game with Obnoxious Effects. In *International Workshop on Frontiers in Algorithmics*. Springer, 172–182.

[30] Reshef Meir. 2019. Strategyproof Facility Location for Three Agents on a Circle. In *SAGT (Lecture Notes in Computer Science, Vol. 11801)*. Springer, Berlin, 18–33.

[31] M.T. Melo, S. Nickel, and F. Saldanha da Gama. 2009. Facility location and supply chain management – A review. *European Journal of Operational Research* 196, 2 (2009), 401–412.

[32] Vijay Menon and Kate Larson. 2019. Mechanism design for locating a facility under partial information. In *Algorithmic Game Theory: 12th International Symposium, SAGT 2019*. Springer, 49–62.

[33] Hervé Moulin. 1980. On strategy-proofness and single peakedness. *Public Choice* 35, 4 (1980), 437–455.

[34] Noam Nisan and Amir Ronen. 1999. Algorithmic mechanism design. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*. ACM, New York, 129–140.

[35] Martin Pal, T Tardos, and Tom Wexler. 2001. Facility location with nonuniform hard capacities. In *Proceedings 42nd IEEE symposium on foundations of computer science*. IEEE, New York, 329–338.

[36] Yves Pochet and Laurence A Wolsey. 1988. Lot-size models with backlogging: Strong reformulations and cutting planes. *Mathematical Programming* 40 (1988), 317–335.

[37] Ariel D Procaccia and Moshe Tennenholtz. 2013. Approximate mechanism design without money. *ACM Transactions on Economics and Computation (TEAC)* 1, 4 (2013), 1–26.

[38] Cédric Villani. 2009. *Optimal transport: old and new*. Vol. 338. Springer, Berlin.

[39] Toby Walsh. 2022. Strategy Proof Mechanisms for Facility Location with Capacity Limits. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. 527–533.

[40] Qiqi Yan. 2011. Mechanism design via correlation gap. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 710–719.

[41] Emmanouil Zampetakis and Fred Zhang. 2023. Bayesian Strategy-Proof Facility Location via Robust Estimation. In *International Conference on Artificial Intelligence and Statistics*. PMLR, London, 4196–4208.

[42] Shaokun Zou and Minming Li. 2015. Facility location games with dual preference. In *Proceedings of the 2015 international conference on autonomous agents and multiagent systems*. ACM, New York, 615–623.