Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Truthful Mechanisms without Money for Non-Utilitarian Heterogeneous Facility Location

Paolo Serafino
School of Computing
Teesside University
p.serafino@tees.ac.uk

Carmine Ventre
School of Computing
Teesside University
c.ventre@tees.ac.uk

Abstract

In this paper, we consider the facility location problem under a novel model recently proposed in the literature, which combines the no-money constraint (i.e. the impossibility to employ monetary transfers between the mechanism and the agents) with the presence of heterogeneous facilities, i.e. facilities serving different purposes. Agents thus have a significantly different cost model w.r.t. the classical model with homogeneous facilities studied in literature. We initiate the study of non-utilitarian optimization functions under this novel model. In particular, we consider the case where the optimization goal consists of minimizing the maximum connection cost of the agents. In this setting, we investigate both deterministic and randomized algorithms and derive both lower and upper bounds regarding the approximability of strategyproof mechanisms.

1 Introduction

This work is motivated by a scenario of big data distribution in clouds. Consider a multinational company having to decide how to distribute the data contained in its databases over its data network. Not all the various offices working for the company need access to the whole data, e.g., a payroll office arguably needs access to employees' data but not the customers', whilst sales offices need customers' data but not employees'. Thus, a demand-based allocation seems a sensible approach. However, beside all the problems that replication might involve, space might be too limited a resource to allow a replication of requested data to all the demanding offices. Fast data access becomes then competitive and, guided by their willingness to have prompt access to the data they need, offices might strategize and amend their demands accordingly. The company, however, wants to minimize the maximum access time in order to guarantee a decent quality of service so that each office can work efficiently.

Mechanism design is the typical answer to situations in which agents can act selfishly and there is a divergence between designer's and agents' objectives. Indeed, we can cast the aforementioned scenario into a mechanism design problem. We are given a graph $G = (V, E)$ representing the data network of the company, wherein the edges represent direct links between data servers. Offices (selfish agents) reside on the nodes, declare what content (e.g., tables) of the database they need to use and seek to minimize their individual connection cost (defined as length of paths in G) to access their data. From the company's perspective, instead, the desideratum is to compute an allocation (i.e., assign each content to a node of the graph) for which the maximum overall connection cost is minimized. The mechanism should then be *truthful*: the strategy maximizing the "happiness" of agents, called *utility*, is *truthtelling*. Truthfulness is typically guaranteed by means of monetary transfers between designer and agents. However, in this setting, it is not conceivable to charge a payroll office because it does not need customer data or to compensate a remote sales office that has to endure a slow access to distantly located customer records. We then want to study truthful (or *strategyproof* (*SP*)) mechanisms without monetary transfers.

Generally speaking, this work belongs to the research agenda of approximate mechanism design without money, recently proposed in (Procaccia and Tennenholtz 2013). In this family of mechanisms, truthfulness is typically enforced by exploiting the *approximation ratio* of the mechanism in those cases where the optimal outcome is not truthful. Our motivating scenario above bears some similarities with the archetypal problem in this agenda, facility location. Our model has however two sources of novelty. With respect to the main stream of works on facility location, we do not consider "homogeneous" (i.e., serving the same purpose) facilities, wherein the cost of an agent is defined as the cost of connecting to the nearest facility, but rather *heterogeneous* (i.e., serving different purposes) facilities, and hence *the cost of an agent is defined as the cost of accessing the facilities she bids for*. Such a model has been introduced very recently in (Serafino and Ventre 2014) where we focused only on the utilitarian objective function of minimizing the total connection cost of the agents (i.e., minimizing the social cost). We differentiate from our previous work in that Min-Max, the objective in the present paper, sits on the other end of the spectrum.

Min-Max is a popular objective function in optimization, mechanism design with (cf., e.g., the rich literature on scheduling selfish machines dating back to (Nisan and Ronen 2001)) and without money (see, e.g., (Koutsoupias 2014)), and studied already for homogeneous facility location (Procaccia and Tennenholtz 2013). In particular, this objective function seems to be rather important to measure *fairness* of allocation in the context of facility location, as envy freeness – the concept usually adopted for fairness – is not applicable when agents cannot exchange outcomes.

Our Contribution

We focus on the heterogeneous facility location problem in the case in which G is a line and we have two facilities to locate, the same setting previously studied in (Serafino and Ventre 2014). Despite its apparent simplicity, this class of instances models the aforementioned content distribution scenario (G being the backbone of the company's data network; facilities being employee and customer records) yet already encodes many intricacies. We study both deterministic and randomized algorithms, and prove that in both cases the optimal allocation does not preserve truthfulness.

We prove a lower bound of $3/2$ on the approximation guarantee of deterministic *SP* mechanisms. The proof connects three different instances and uses truthfulness constraints on two agents to establish the lower bound. This is somehow more complex than typical lower bounds in literature wherein two instances and one lying agent are normally considered. As we already noted in (Serafino and Ventre 2014), designing deterministic *SP* mechanisms for the problem appears difficult. We then analyze a deterministic *SP* algorithm for the location of heterogeneous facility already introduced in the literature (Serafino and Ventre 2014), and prove it is 3–approximate in our scenario. We observe that it is also *SP* (as strategyproofness is independent from the objective function of the mechanism but depends solely on the agents' cost function), and constitutes a tight approximation guarantee for our problem.

Regarding randomized mechanisms we first prove a lower bound of $4/3$ and then design a $3/2$-approximate randomized *SP* mechanism. This algorithm is mainly based on the idea of allocating (in expectation) each facility on the average position of the subgraph of G comprised of agents requesting it. This way truthfulness is guaranteed since there is no advantage in hiding one's own requested facilities as the aforementioned subgraph can only move away from the lying agent. (Note that adding unneeded facilities does not help either.) A complication to this intuition is that facilities cannot always be located in the "middle" of the subgraph. Our algorithm works around this, while preserving truthfulness and guaranteeing a good approximation.

Roadmap. The remainder of this paper is organized as follows. Section 2 briefly surveys some related literature. In Section 3 we formalize our model for the heterogeneous facility location problem on the line. In Section 4 we discuss our results about deterministic algorithms whereas in Section 5 we present our results for randomized algorithms. Finally, in Section 6 we draw some conclusions and highlight some future research efforts.

2 Related Work

The facility location problem has been extensively studied from a Mechanism Design perspective under both utilitarian and non-utilitarian objective functions. We briefly review some literature in both scenarios.

As for utilitarian objective functions, in (Moulin 1980) the author studies the case of a single homogeneous facility on the line and *single-peaked* agents (i.e., agents whose valuation has a single peak – their preferred location – and decreases as the location of the facility gets farther from the peak) and proves that generalized median voter schemes are the only deterministic *SP* mechanism. Schummer and Vohra (Schummer and Vohra 2002) extend this result to *continuous graphs*, characterizing truthfulness on *continuous lines and trees*. They show that on circular graphs every *SP* mechanism must be dictatorial.

In the aforementioned paper (Procaccia and Tennenholtz 2013) the authors initiated the field of approximate mechanism design without money by studying the problems of truthfully locating (without money) one and two homogeneous facilities, wherein agents can lie about their location on a continuous line. They focus on both social cost and min-max objective functions. For 2-facility location and utilitarian objective, they propose the Two-Extremes algorithm, that places the two facilities in the leftmost and rightmost location of the instance, and prove that it is group strategyproof and has a linear (in the number of agents) approximation. This lower bound has later been shown to be tight by the characterization of truthfulness given in (Fotakis and Tzamos 2013): the authors show that there are only two deterministic *SP* mechanisms with *bounded approximation ratio* for the 2-facility location problem on the line: Dictatorial and Two-Extremes. Lu et al. (Lu, Wang, and Zhou 2009) prove a 1.045 lower bound for randomized mechanisms and an $n/2$-approximate lower bound for randomized mechanisms, thus improving the bounds given in (Procaccia and Tennenholtz 2013). Alon et al. (Alon et al. 2010) study *continuous cycles* and derive a linear (in the number of agents) lower bound for *SP* mechanisms and a constant approximation bound for randomized mechanisms. Closer to our setting is the work by Dokow et al (Dokow et al. 2012) who focus on *discrete* lines and cycles instead, and prove that *SP* mechanisms on *discrete large cycles* are nearly-dictatorial, since all agents are able to effect the outcome to a certain (albeit limited) extent. Contrarily to the case of continuous cycles studied in (Schummer and Vohra 2002), for small discrete graphs Dokow et al. prove that there are *anonymous* (i.e., independent from agents' IDs) *SP* mechanisms. Furthermore, they prove a linear lower bound in the number of agents for the approximation ratio of *SP* mechanisms on discrete cycles. The model of heterogeneous 2-facility location we study in this paper has been introduced in (Serafino and Ventre 2014), where we studied both deterministic and randomized strategyproof algorithms, with the aim to minimize the social cost. Furthermore, we proved an approximation lower bound of $9/8$ for deterministic algorithms, and proved that a simple adaptation of the Two-Extremes algorithm proposed in (Procaccia and Tennenholtz 2013) provides a linear approximation, in the number of agents, to the optimum. Interestingly, we proved that the optimum is strategyproof in expectation.

The literature on Min-Max objective function is quite rich in the case of mechanism design with money (mainly because it shows the tension between approximation and truthfulness – being VCG not applicable) but sparse in the case of moneyless mechanisms. Procaccia and Tennenholtz prove in their model tight bounds for min-max approximation with 1 facility and nearly tight results with 2 facilities. The paper (Koutsoupias 2014) studies moneyless *SP* mechanisms approximating min-max objective for scheduling selfish unrelated machines whose execution times can be verified.

3 Maximum Cost Model

We study the min-max objective function for the *heterogeneous 2-facility location problem on the line* (hereinafter facility location, for short). We need to locate two facilities on a *linear unweighted graph* so as to minimize the maximum connection cost of the agents.

The input to the problem consists of a set of agents $N = \{1, \ldots, n\}$; an undirected unweighted linear graph $G = (V, E)$, where $V \supseteq N$; and a set of facilities $\mathfrak{F} = \{\mathsf{F}_1, \mathsf{F}_2\}$. The *types* of the agents are subsets of $\mathfrak{F}$, called their *facility set*. We let $T_i \subseteq \mathfrak{F}$ denote the true type of agent i. Sometimes, slightly abusing notation, we will regard T_i as a set of *indices* j s.t. $\mathsf{F}_j \in T_i$. A mechanism M for the facility location problem takes as input a vector of types $\mathcal{T} = (T_1, \ldots, T_n)$ and returns as output a feasible allocation $M(\mathcal{T}) = (F_1, F_2)$, such that $F_i \in V$, $i = 1, 2$, and $F_1 \neq F_2$. Given a feasible allocation $\mathcal{F} = (F_1, F_2)$, agent i has a cost defined as $cost_i(\mathcal{F}) = \sum_{j \in T_i} d(i, F_j)$, where $d(i, F_j)$ denotes the length of the shortest path from i to F_j in G. Agents seek to minimize their cost and could, therefore, misreport their facility sets to the mechanism if this reduces their cost. We let $T'_i \subseteq \mathfrak{F}$ denote a declaration of agent i to the mechanism. We are interested in the following class of mechanisms.

Definition 3.1. *A mechanism M is* truthful *(or* strategyproof, SP*, for short) if for any i, any declaration T'_i of agent i, any declarations of the other agents $\mathcal{T}_{-i}$ we have $cost_i(\mathcal{F}) \leq cost_i(\mathcal{F}')$, where $\mathcal{F} = M(\mathcal{T})$ and $\mathcal{F}' = M(T'_i, \mathcal{T}_{-i})$. A randomized M is* truthful in expectation *if the* expected cost *of every agent is minimized by truthtelling.*

We want truthful mechanisms M that return an allocation $\mathcal{F} = M(\mathcal{T})$ that minimizes the *maximum cost* function $mc(\mathcal{F}) = \max_{i \in N} cost_i(\mathcal{F})$, namely: $M(\mathcal{T}) \in \operatorname{argmin}_{\mathcal{F} \text{ feasible}} mc(\mathcal{F})$. We call these mechanisms *optimal* and denote an optimal allocation on declaration vector $\mathcal{T}$ as $OPT(\mathcal{T})$ if $mc(OPT(\mathcal{T})) = \min_{\mathcal{F} \text{ feasible}} mc(\mathcal{F})$. Sometimes truthfulness and optimality are incompatible and therefore we have to content ourselves with sub-optimal solutions. In particular, we say that a mechanism M is α-approximate if $mc(M(\mathcal{T})) \leq \alpha \cdot mc(OPT(\mathcal{T}))$. Furthermore, we denote as $V_j[\mathcal{T}]$ the set of agents wanting access to facility F_j according to a declaration vector $\mathcal{T}$, i.e., $V_j[\mathcal{T}] = \{i \in N | \mathsf{F}_j \in T_i\}$.

For the sake of notational conciseness, in the remainder of the paper we will often omit the declaration vector $\mathcal{T}$ (e.g., $V_k[\mathcal{T}]$ simply denoted as V_k) and denote an untruthful declaration $(T'_i, \mathcal{T}_{-i})$ of agent i by a prime symbol (e.g., $V_k[T'_i, \mathcal{T}_{-i}]$ simply denoted as V'_k).

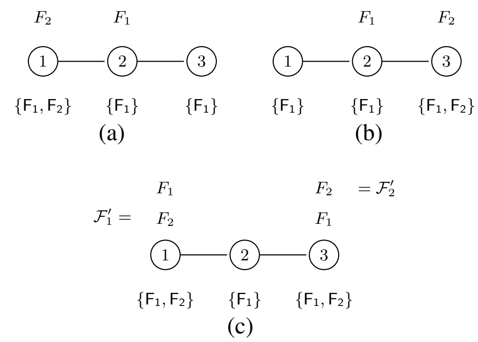

Figure 1: Instances used to prove the lower bound of $3/2$

4 Deterministic Mechanisms

In this section we analyze deterministic mechanisms for the min-max heterogeneous facility location problem.

Lower bound

We start by presenting a negative result stating the impossibility of approximating the optimal allocation within 3/2 of the optimal value while maintaining strategyproofness.

Theorem 4.1. *There exists no α-approximate deterministic SP algorithm for the facility location problem with $\alpha < 3/2$.*

Proof. Let us first consider the two instances depicted in Figure 1(a) and Figure 1(b). The agents in the instance of Figure 1(a) have declarations $T_1 = \{\mathsf{F}_1, \mathsf{F}_2\}$, $T_2 = \{\mathsf{F}_1\}$ and $T_3 = \{\mathsf{F}_1\}$, whereas the agents in Figure 1(b) have declarations $T_1 = \{\mathsf{F}_1\}$, $T_2 = \{\mathsf{F}_1\}$ and $T_3 = \{\mathsf{F}_1, \mathsf{F}_2\}$. It is easy to check that the optimal allocation for the instance of Figure 1(a) is $\mathcal{F}_1 = (F_1 = 2, F_2 = 1)$, whereas the optimal allocation for the instance of Figure 1(b) is $\mathcal{F}_2 = (F_1 = 2, F_2 = 3)$, having both cost 1. We note that, in both cases, any second-best solution has cost 2, so any $3/2$-approximate algorithm would return the optimal solution for these instances. Let us now consider the case when agent 3 in instance 1(a) lies declaring $T'_3 = \{\mathsf{F}_1, \mathsf{F}_2\}$ (see Figure 1(c)). In this case we have two optimal solutions: $\mathcal{F}'_1 = (F_1 = 3, F_2 = 1)$ and $\mathcal{F}'_2 = (F_1 = 1, F_2 = 3)$, with cost 2. We note that we obtain the same instance (and hence the same optimal solutions) if we consider the case when agent 1 in instance 1(b) lies declaring $T'_1 = \{\mathsf{F}_1, \mathsf{F}_2\}$ instead of her true type. We now note that neither $\mathcal{F}'_1$ nor $\mathcal{F}'_2$ are *SP*. In fact, if $\mathcal{F}'_1$ is returned, we can then regard the instance of Figure 1(c) as resulting from the instance of Figure 1(a) when agent 3 lies, in which case agent 3 would gain by lying, as $cost_3(\mathcal{F}_1) = 1 > 0 = cost_3(\mathcal{F}'_1)$. On the other hand, if $\mathcal{F}'_2$ is returned, we can then regard the instance of Figure 1(c) as resulting from the instance of Figure 1(b) when agent 1 lies, in which case agent 1 would gain by lying, as $cost_1(\mathcal{F}_2) = 1 > 0 = cost_1(\mathcal{F'}_2)$. It is clear from the above argument that an *SP* solution for the instance of Figure 1(c) locates facility F_1 at node 2, hence the only *SP* solutions are $(F_1 = 2, F_2 = 1)$ and $(F_1 = 2, F_2 = 3)$. Since the cost of these solutions is 3, the claim is proven. □

Since we have proved a $3/2$ approximation lower bound for deterministic SP algorithms, the following corollary easily follows.

Corollary 4.2. *There is no optimal deterministic SP algorithm for the facility location problem.*

Upper bound

We now discuss a 3-approximate *SP* algorithm, named TWOEXTREMES, given in (Serafino and Ventre 2014); the algorithm is inspired by (Procaccia and Tennenholtz 2013).

Algorithm 1: TWOEXTREMES

Require: Line G, facilities $\mathfrak{F} = \{\mathsf{F}_1, \mathsf{F}_2\}$, declarations $\mathcal{T} = \{T_1, \dots, T_n\}$
Ensure: $F(\mathcal{T})$, a 3-approximate allocation for 2-facility location on G

```
1: F_1 := min V_1[T]
2: F_2 := max V_2[T]
3: if F_1 = F_2 then
4:     if F_2 − 1 ≠ NIL then
5:         F_2 := F_2 − 1
6:     else
7:         F_1 := F_1 + 1
8:     end if
9: end if
10: return (F_1, F_2)
```

TWOEXTREMES locates facility F_1 in the leftmost location of $V_1[\mathcal{T}]$ and F_2 in the rightmost location of $V_2[\mathcal{T}]$. If a tie occurs (line 3) the algorithm considers whether the chosen location is *not* the leftmost location of the instance, in which case F_2 is moved left on the nearest node; otherwise, F_1 is moved right on the nearest node. This tie breaking rule is essential to prove the strategyproofness of the algorithm. Below, we state and prove for completeness this fact.

Theorem 4.3. *(Serafino and Ventre 2014) Algorithm* TWOEXTREMES *is SP.*

Proof. For the sake of contradiction, let us assume that there exist $i \in N$ with type T_i and an untruthful declaration T_i' such that $\sum_{j \in T_i} d(i, F_j(\mathcal{T})) > \sum_{j \in T_i} d(i, F_j(T_i', \mathcal{T}_{-i}))$, where $F_j(\mathcal{Z})$ denotes the location to which TWOEXTREMES, on input the declaration vector $\mathcal{Z}$, assigns facility F_j. We need to analyse three cases: (*a*) $i = \min V_1$, (*b*) $i = \max V_2$, and (*c*) $i \notin \{\min V_1[\mathcal{T}], \max V_2[\mathcal{T}]\}$.

If case (*a*) occurs, it can be either $T_i = \{\mathsf{F}_1\}$ or $T_i = \{\mathsf{F}_1, \mathsf{F}_2\}$. If $T_i = \{\mathsf{F}_1\}$ then $F_1 = i$, $cost_i(F(\mathcal{T})) = 0$ and i cannot decrease her cost any further by misreporting her type. If $T_i = \{\mathsf{F}_1, \mathsf{F}_2\}$, then it can be either $i = \max V_2$ (in which case the algorithm returns $(F_1 = i-1, F_2 = i)$ or $(F_1 = i, F_2 = i+1)$, $cost_i(\mathcal{F}) = 1$ and i cannot decrease her cost any further by lying) or $i < \max V_2$ (in which case $F_1 = i$ and i cannot influence the location of facility F_2).

It is easy to check that case (*b*) is symmetric to case (*a*).

If case (*c*) occurs, then it can be either: $T_i = \{\mathsf{F}_1\}$, $T_i = \{\mathsf{F}_2\}$ or $T_i = \{\mathsf{F}_1, \mathsf{F}_2\}$. If $T_i = \{\mathsf{F}_1\}$, then $i > \min V_1$. It is easy to check that if $\min V_1 \neq \max V_2$ then i cannot influence the location of facility F_1. Let us assume then that $\ell = \min V_1 = \max V_2$. In this case the algorithm outputs either $(F_1 = \ell, F_2 = \ell - 1)$ or $(F_1 = \ell + 1, F_2 = \ell)$. In either case, if $T_i' = \emptyset$ the output of the algorithm does not change, whereas if $\mathsf{F}_2 \in T_i'$ then the algorithm outputs $(F_1' = \ell, F_2' = i)$ (as $i > \max V_2$) and $cost_i(F(\mathcal{T})) \leq cost_i(F(T_i', \mathcal{T}_i))$. It is easy to check that the case when $T_i = \{\mathsf{F}_2\}$ is symmetric to the case when $T_i = \{\mathsf{F}_1\}$.

If $T_i = \{\mathsf{F}_1, \mathsf{F}_2\}$ then $\min V_1 < i < \max V_2$, and it is easy to check that i cannot influence the outcome of the algorithm. □

We now prove that TWOEXTREMES is 3-approximate and show that our analysis is tight.

Theorem 4.4. *The* TWOEXTREMES *algorithm is 3-approximate.*

Proof. Let $\mathcal{F}^*$ denote the optimal allocation and $\mathcal{F}$ denote the allocation returned by the TWOEXTREMES algorithm. Let us consider a node i such that $i \in \operatorname{argmax}_{i \in N} cost_i(\mathcal{F})$ and let us denote $EXT = cost_i(\mathcal{F})$. It is easy to check that the following holds:

$$cost_i(\mathcal{F}^*) = EXT - \Delta F \leq OPT \tag{1}$$

where $\Delta F = \sum_{j \in T_i} \Delta F_j$ and $\Delta F_j = d(i, F_j) - d(i, F_j^*)$. Intuitively, (1) formalizes the fact that the optimal allocation locates the facilities closer to i with respect to TWOEXTREMES in order to lower the cost of agent i. Let S denote the set $\{\min V_1, \max V_2\}$ if $T_i = \{\mathsf{F}_1, \mathsf{F}_2\}$, $\{\min V_1\}$ if $T_i = \{\mathsf{F}_1\}$ and $\{\max V_2\}$ if $T_i = \{\mathsf{F}_2\}$. Because of the change of facilities' position, there is an agent $x \in S$ and a facility $k \in T_i \cap T_x$ such that $d(x, F_k) \leq d(x, F_k^*)$. It is not hard to check that the following holds for $x \in S$:

$$\begin{aligned} OPT \geq \quad & cost_x(\mathcal{F}^*) \geq d(x, F_k^*) \geq \\ \geq \quad & d(x, F_k^*) - d(x, F_k) \geq \\ \geq \quad & d(i, F_k) - d(i, F_k^*) \geq \quad \Delta F_k. \end{aligned} \tag{2}$$

Two cases can occur: (*i*) $|T_i| = 1$ and (*ii*) $|T_i| = 2$. If case (*i*) occurs, we notice that $\Delta F = \Delta F_k$, $k \in T_i$, and from (1) and (2) we derive $EXT \leq 2 \cdot OPT$. If case (*ii*) occurs, we notice that applying (2) with k and $k+1$, we obtain $2 \cdot OPT \geq \Delta F_k + \Delta F_{k+1}$, and, finally, from (1), $EXT \leq 3 \cdot OPT$. □

Theorem 4.5. *The upper bound of Theorem 4.4 is tight.*

Proof. We are now going to prove that the bound is tight. Let us consider the 3-agent family of instances depicted in Figure 2, where between nodes 1 and 2 and between nodes 4 and 5 there are λ "empty" nodes whereas between nodes 2 and 3 and between nodes 3 and 4 there are $\frac{\lambda}{2}$ "empty" nodes. It is easy to check that the optimal allocation in this case is $\mathcal{F}^* = (F_1^* = 2, F_2^* = 4)$ and the optimal cost is $mc(\mathcal{F}^*) = \lambda$. Furthermore, it is easy to check that the cost of the allocation $\mathcal{F} = (F_1 = 1, F_2 = 5)$ computed by algorithm TWOEXTREMES is $mc(\mathcal{F}) = cost_3(\mathcal{F}) = 3\lambda$, which is indeed 3-approximate. □

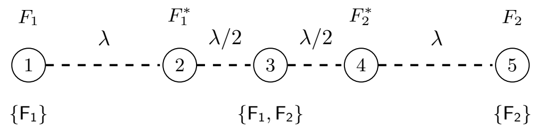

Figure 2: Instance showing that the bound of Theorem 4.4 is tight.

5 Randomized Mechanisms

In this section we shift our focus on randomized algorithms. Similarly to the case of deterministic mechanisms, our results are twofold: (*i*) we prove a negative result regarding the impossibility of obtaining arbitrarily good approximations by means of *SP* algorithms and (*ii*) we present a $3/2$-approximate randomized *SP* algorithm.

Lower bound

We begin by proving the impossibility of approximating the optimum within a factor of $4/3$ and preserve the strategyproofness.

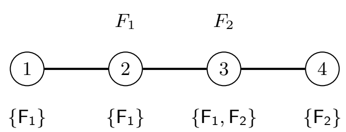

Figure 3: Truthful instance used to prove the bound of Theorem 5.1

Theorem 5.1. *There is no SP randomized α-approximate algorithm with $\alpha < 4/3$.*

Proof. Let us consider the instance depicted in Figure 3. It is easy to check that the (unique) optimal solution $\mathcal{F}^* = (F_1 = 2, F_2 = 3)$ has cost $mc(\mathcal{F}^*) = 1$, whereas any suboptimal allocation has cost at least 2. Let us consider a generic randomized algorithm A that returns the optimal allocation with probability ρ and some suboptimal allocations with probability $1 - \rho$. If A is $4/3$-approximate, then $4/3 \geq mc(A(\mathcal{T})) \geq 1 \cdot \rho + 2(1 - \rho)$ which implies that A must return the exact solution with probability $\rho \geq 2/3$. In particular, this means that for agent 4 the allocation computed by A has cost $cost_4(A(\mathcal{T})) \geq 2/3$. If agent 4 lies declaring $T_4' = \{\mathsf{F}_1, \mathsf{F}_2\}$ (see Figure 4), then the optimal allocations are $\mathcal{F}_1' = (F_1 = 2, F_2 = 4)$ and $\mathcal{F}_2' = (F_1 = 3, F_2 = 4)$, having cost 2, whereas any suboptimal solution has cost at least 3. Once again, A would return optimal solutions with probability π and suboptimal solutions with probability $1 - \pi$. In particular, to preserve strategyproofness $cost_4(A(\mathcal{T}')) \geq 2/3$ must hold, which implies that $1 - \pi \geq 2/3$ (as both the optimal solutions, $\mathcal{F}_1'$ and $\mathcal{F}_2'$, locate F_2 on 4) and $\pi \leq 1/3$. Hence, $mc(A(\mathcal{T}')) \geq 2 \cdot 1/3 + 3 \cdot 2/3 = 8/3$, which yields an approximation ratio of $\alpha \geq 4/3$. □

Since we have proved a $4/3$ approximation lower bound for randomized SP algorithms, the following corollary easily follows.

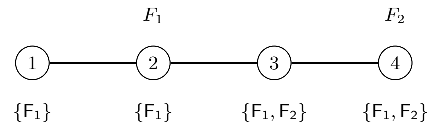

Figure 4: Instance used in the proof of Theorem 5.1 wherein agent 4 lies

Corollary 5.2. *There is no optimal randomized SP algorithm for the facility location problem.*

Upper bound

We now present a randomized *SP* algorithm, which returns $3/2$-approximate solutions. The main idea behind the algorithm, called RANDAVG and presented in Algorithm 2, is to locate in expectation facility F_k on the mean location of V_k, thus guaranteeing that hiding $\mathsf{F_k}$ from one's own type is not profitable (i.e., V_k can only shrink away from the lying agent). The algorithm uses a procedure COMPUTESUPPORT(x, y) that returns a set of deterministic feasible solutions called *mean set*. The mean set returned by procedure COMPUTESUPPORT(x, y) is such that a feasible solution extracted uniformly at random from it has the property that the expected location of F_0 (F_1, respectively) is on x (y, respectively).[1] There are, however, certain extreme situations in which the existence of mean sets is not guaranteed (cf. Lemma 5.3). RANDAVG needs to consider these cases separately (cf. lines 3 and 6 of the algorithm) and return deterministic solutions instead.

In this section, we let μ_k denote the average location of V_k, i.e., $\mu_k = \text{avg}(V_k) = \frac{\min V_k + \max V_k}{2}$, $k = 0, 1$. Depending on the parity of $|V_k|$, μ_k might either lie on a vertex of G (if $|V_k|$ is odd) or in between two vertices (if $|V_k|$ is even); we denote the former case as $\mu_k \in V$ and the latter as $\mu_k \in E$ (meaning that μ_k lies on an edge of G, formally: $\exists(u, v) \in E$ such that $\mu_k = (u + v)/2$). We let RIGHT$(\mu_k) = \lceil \mu_k \rceil$ if $\mu_k \in E$ and $\mu_k + 1$ otherwise. Similarly, LEFT$(\mu_k) = \lfloor \mu_k \rfloor$ if $\mu_k \in E$ and $\mu_k - 1$ otherwise.

Lemma 5.3. *Let $G = (V, E)$ be the network on which agents reside. There always exists a mean set for graph G if either of the following holds:* (i) $|\mu_k - \mu_{k+1}| \geq 1$, (ii) $\forall k \in \{0, 1\}$, RIGHT$(\mu_k) \neq$ NIL *and* LEFT$(\mu_k) \neq$ NIL.

Proof. Let us focus on case (*i*) initially. We distinguish the cases in which μ_k is in V from those in which it is in E. If $\mu_k \in V$ and $\mu_{k+1} \in V$ then $\mathcal{M} = \{(\mu_k, \mu_{k+1})\}$ is a mean set for G (note that this solution is feasible as $|\mu_k - \mu_{k+1}| \geq 1$). If $\mu_k \in E$ and $\mu_{k+1} \in V$ then: LEFT$(\mu_k) \neq$ NIL, RIGHT$(\mu_k) \neq$ NIL and both LEFT$(\mu_k) \neq \mu_{k+1}$ and RIGHT$(\mu_k) \neq \mu_{k+1}$ (as, by hypothesis, $|\mu_k - \mu_{k+1}| \geq 1$). Hence, $\mathcal{M} = \{(\text{LEFT}(\mu_k), \mu_{k+1}), (\text{RIGHT}(\mu_k), \mu_{k+1})\}$ is a feasible mean set for G. If both $\mu_k \in E$ and $\mu_{k+1} \in E$, then since $|\mu_k - \mu_{k+1}| \geq 1$, we have that $\mathcal{M} = \{(\text{LEFT}(\mu_k), \text{LEFT}(\mu_{k+1})), (\text{RIGHT}(\mu_k), \text{RIGHT}(\mu_{k+1}))\}$ is a mean set for G.

[1] To ease the notation, in this section we use binary indexes for the facilities.

Let us now focus on case (*ii*). We assume that $|\mu_k - \mu_{k+1}| < 1$ (for otherwise the arguments above apply). It is easy to check that the only case that can occur is $\mu_k \in V$ and $\mu_{k+1} \in E$ for some $k \in \{0,1\}$. Then $\mathcal{M} = \{(\textsc{Left}(\mu_k), \textsc{Right}(\mu_{k+1})), (\textsc{Right}(\mu_k), \textsc{Left}(\mu_{k+1}))\}$ is a feasible mean set for G. $\square$

Observe that the proof of the lemma gives a constructive way to implement (in polynomial-time) procedure COMPUTESUPPORT.

Algorithm 2: RANDAVG

Require: Line G, facilities $\mathfrak{F} = \{\mathsf{F}_1, \mathsf{F}_2\}$, declarations $\mathcal{T} = \{T_1, \dots, T_n\}$
Ensure: $F_{AVG}(\mathcal{T})$, a $3/2$-approximate allocation for 2-facility location on G

1: $\mu_k := \text{avg}(V_k), \ \ \forall k \in \{0,1\}$
2: **if** $\exists k \in \{0,1\}$ s.t. $\textsc{Right}(\mu_k) = \textsc{Nil}$ AND $|\mu_0 - \mu_1| < 1$ **then**
3:    **return** $\left(\bar{F}_k := \mu_k, \bar{F}_{k+1} := \textsc{Left}(\mu_{k+1})\right)$
4: **end if**
5: **if** $\exists k \in \{0,1\}$ s.t. $\textsc{Left}(\mu_k) = \textsc{Nil}$ AND $|\mu_0 - \mu_1| < 1$ **then**
6:    **return** $\left(\bar{F}_k := \mu_k, \bar{F}_{k+1} := \textsc{Right}(\mu_{k+1})\right)$
7: **end if**
8: $\mathcal{M} := \textsc{ComputeSupport}(\mu_k, \mu_{k+1})$
9: **return** $(F_k, F_{k+1}) \in \mathcal{M}$ with probability $1/|\mathcal{M}|$

Theorem 5.4. *Algorithm* RANDAVG *is SP in expectation.*

Proof. It is easy to check that algorithm RANDAVG returns either a feasible mean set solution (i.e., line 9) or a feasible deterministic solution (lines 3 and 6) when no mean set solution exists. For the sake of notation, in the remainder we will denote a mean set solution as $\mathcal{M}$ and a deterministic solution as $\mathcal{D}$. Let us denote by i the lying agent; we shall prove that $cost_i(\mathcal{F}') \geq cost_i(\mathcal{F})$, where $\mathcal{F}$ and $\mathcal{F}'$ denote the outcomes of RANDAVG on input the true type of i and a misreport, respectively. (The value of $cost_i(\cdot)$ must be intended here with respect to the expected locations of the facilities in T_i.) The analysis distinguishes what type of allocation (i.e., $\mathcal{M}$ or $\mathcal{D}$) $\mathcal{F}$ and $\mathcal{F}'$ are. By letting $\mathcal{X} \to \mathcal{Y}$ symbolize that $\mathcal{F}$ is of type $\mathcal{X}$ and $\mathcal{F}'$ is of type $\mathcal{Y}$, we consider three cases: (*a*) $\mathcal{M} \to \mathcal{M}$; (*b*) $\mathcal{M} \to \mathcal{D}$; (*c*) $\mathcal{F}$ is of type $\mathcal{D}$.

Case (*a*). Let $i \in \{\min V_1, \max V_1, \min V_2, \max V_2\}$. In this case each facility is located independently (in expectation) and truthfulness follows from the simple observation that $d(i, \text{avg}(V_k)) \leq d(i, \text{avg}(V_k'))$, for $k \in T_i$. If $i \notin \{\min V_1, \max V_1, \min V_2, \max V_2\}$, it is easy to check that i cannot alter the outcome of RANDAVG.

Case (*b*). Let us consider the case $\mathcal{M} \to \mathcal{D}$. Since $\mathcal{F}'$ is of type $\mathcal{D}$, $|\mu_0' - \mu_1'| < 1$ and there exists $k \in \{0,1\}$ such that either $\textsc{Left}(\mu_k') = \textsc{Nil}$ or $\textsc{Right}(\mu_k') = \textsc{Nil}$. We focus on the former case; the other case follows by symmetry. Since $\textsc{Left}(\mu_k') = \textsc{Nil}$ then, by definition of LEFT, $\mu_k' \in V$ and, in turns, by definition of μ_k', $V_k' = \{\ell\}$ and $\mu_k' = \ell$. Note also that since $|\mu_k' - \mu_{k+1}'| < 1$ then $\mu_{k+1}' \in E$, with $\textsc{Left}(\mu_{k+1}') = \ell$. In this case we then have $T_\ell' = \{\mathsf{F_k}, \mathsf{F_{k+1}}\}$, $T_{\ell+1}' = \{\mathsf{F_{k+1}}\}$ and $T_r' = \emptyset$ for all $r \geq \ell + 2$, but $T_l = T_l'$ for all $l \neq i$ (i.e. only agent i is lying). On this instance RANDAVG returns $\mathcal{F}' = (F_k' = \ell, F_{k+1}' = \ell + 1)$. Let us assume that $i = \ell$. In this case, if $T_i = \{\mathsf{F_k}\}$, then $\mathcal{F} = (F_k = \ell, F_{k+1} = \ell + 1)$ and the cost of agent i is unchanged, whereas if $T_i = \{\mathsf{F_{k+1}}\}$ then $F_{k+1} = \ell + 0.5$ and $cost_i(\mathcal{F}') = 1 > 0.5 = cost_i(\mathcal{F})$. Let us assume that $i = \ell + 1$. We need to consider two cases: $T_i = \{\mathsf{F_k}\}$ and $T_i = \{\mathsf{F_k}, \mathsf{F_{k+1}}\}$. If $T_i = \{\mathsf{F_k}\}$, then RANDAVG outputs $\mathcal{F} = (F_k = \ell + 1, F_{k+1} = \ell)$ and $cost_i(\mathcal{F}) = 0 < 1 = cost_i(\mathcal{F}')$. If $T_i = \{\mathsf{F_k}, \mathsf{F_{k+1}}\}$, then $\mathcal{F} = (F_k = F_{k+1} = \ell + 0.5)$ and $cost_i(\mathcal{F}) = 1 = cost_i(\mathcal{F}')$. Let us assume that $i \geq \ell + 2$. In this case it can be $T_i = \{\mathsf{F_k}\}$, $T_i = \{\mathsf{F_{k+1}}\}$ or $T_i = \{\mathsf{F_k}, \mathsf{F_{k+1}}\}$. It is easy to check that in all three cases $cost_i(\mathcal{F}') \geq cost_i(\mathcal{F})$.

Case (*c*). Since $\mathcal{F}$ is of type $\mathcal{D}$, by the same reasoning of case (*b*) we conclude that there exists $k \in \{0,1\}$ such that $V_k = \{\ell\}$, $\mu_k = \ell$ and either $\textsc{Left}(\mu_k) = \textsc{Nil}$ or $\textsc{Right}(\mu_k) = \textsc{Nil}$. We focus on the former case, the other begin symmetric. We note that in this case we have a 2-agents instance such that $T_\ell = \{\mathsf{F_k}, \mathsf{F_{k+1}}\}$, and $T_{\ell+1} = \{\mathsf{F_{k+1}}\}$. Since the allocation outputted by RANDAVG is $(F_k = \ell, F_{k+1} = \ell + 1)$ it is easy to check that no agent can lower her cost any further. $\square$

Theorem 5.5. *Algorithm* RANDAVG *is* $3/2$*-approximate.*

Proof. We note that whenever algorithm RANDAVG returns a deterministic solution, then it returns an optimal allocation. Hence, in the remainder we restrict ourselves to considering only the case when RANDAVG returns a mean set solution.

Let us denote a *bottleneck agent* (i.e., an agent incurring the maximum cost) of RANDAVG by i, namely: $cost_i(\mathcal{F}) = mc(\mathcal{F})$, where $\mathcal{F}$ denotes the output of RANDAVG. (Again, $cost_i(\cdot)$ must be considered w.r.t. the expected locations of the facilities in T_i.) Hereinafter, $\mathcal{F}^*$ will denote the optimal solution. We can assume that $T_i = \{\mathsf{F}_0, \mathsf{F}_1\}$, as otherwise RANDAVG would return an optimal allocation. In details, by letting $T_i = \{k\}$, we have $AVG = mc(\mathcal{F}) = \frac{R_k - L_k}{2} \leq mc(\mathcal{F}^*) = OPT$, where $L_k = \min V_k$ and $R_k = \max V_k$.

Let us denote $\Delta F_j = d(i, F_j) - d(i, F_j^*)$, for $j \in \{0,1\}$. It is easy to check that:

$$cost_i(\mathcal{F}^*) = AVG - \Delta F \leq OPT, \tag{3}$$

where $\Delta F = \Delta F_0 + \Delta F_1$. Intuitively, the optimal allocation locates the facilities closer to i with respect to RANDAVG in order to lower the cost of agent i. Because of this, there is an agent $x \in \{L_0, R_0\}$ such that $d(x, F_0) \leq d(x, F_0^*)$. It is not too hard to check that the following holds:

$$OPT \geq cost_x(\mathcal{F}^*) \geq \frac{R_0 - L_0}{2} + \Delta F_0. \tag{4}$$

Likewise, there is an agent $y \in \{L_1, R_1\}$ such that $d(y, F_1) \leq d(y, F_1^*)$ and we have:

$$OPT \geq cost_y(\mathcal{F}^*) \geq \frac{R_1 - L_1}{2} + \Delta F_1. \tag{5}$$

We now need to consider two cases: $\Delta F \leq \frac{OPT}{2}$ and $\Delta F > \frac{OPT}{2}$. If $\Delta F \leq \frac{OPT}{2}$, the claim follows immediately from

(3), as $AVG - \frac{OPT}{2} \leq AVG - \Delta F$ holds. If $\Delta F > \frac{OPT}{2}$, then the following holds:

$$\begin{aligned}\frac{OPT}{2} &< \Delta F_0 + \Delta F_1 \\ &\leq 2 \cdot OPT - \frac{R_0 - L_0}{2} - \frac{R_1 - L_1}{2}\end{aligned} \tag{6}$$

where the second inequality follows from (4) and (5). From (6), we obtain

$$\frac{R_0 - L_0}{2} + \frac{R_1 - L_1}{2} < \frac{3}{2} \cdot OPT.$$

By observing that

$$AVG \leq \frac{R_0 - L_0}{2} + \frac{R_1 - L_1}{2},$$

the claim follows. □

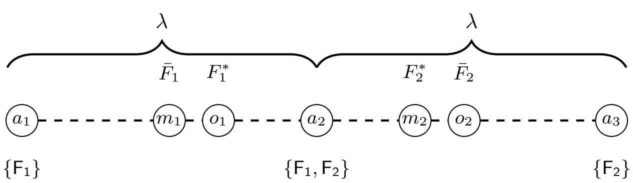

Figure 5: Tight instance for RANDAVG

Theorem 5.6. *The upper bound of Theorem 5.5 is tight.*

Proof. Figure 5 depicts a family of instances for which the RANDAVG algorithm always returns a $3/2$-approximate solution, thus showing that the analysis above is tight. This family of instances consists of (at least) 3 agents a_1, a_2 and a_3 such that: *(i)* $a_1 < a_2 < a_3$; *(ii)* $d(a_1, a_2) = d(a_2, a_3) = \lambda$, where $\lambda \in \mathbb{Z}$; *(iii)* $a_1 = \min V_1$, $a_2 = \max V_1 = \min V_2$ and $a_3 = \max V_2$; *(iv)* $T_a = \{\mathsf{F}_1\}$, $T_b = \{\mathsf{F}_1, \mathsf{F}_2\}$ and $T_c = \{\mathsf{F}_2\}$. It is easy to check that, for this family of instances, the optimal allocation is $(F_1^* = o_1, F_2^* = o_2)$, such that $o_1 = \frac{2}{3}\lambda$ and $o_2 = \frac{4}{3}\lambda$, and $mc((F_1^*, F_2^*)) = \frac{2}{3}\lambda$. Algorithm RANDAVG returns allocation $(\bar{F}_1 = m_1, \bar{F}_2 = m_2)$ such that: $m_1 = \frac{a_1+a_2}{2}$, $m_2 = \frac{a_2+a_3}{2}$ and $mc((\bar{F}_1, \bar{F}_2)) = \lambda$. Hence, the approximation ratio of algorithm RANDAVG on this family of instances is $3/2$. □

6 Conclusions

We have studied the heterogeneous facility location problem with a non-utilitarian optimization function, namely the maximum connection cost of the agents. More in general, works falling in this research agenda often deal only with single-parameter agents (exceptions being the studies on mechanisms without money and verification (Koutsoupias 2014; Fotakis, Krysta, and Ventre 2014)).

We have shown that even for very simple agents' domains, comprised of only 2 bits, truthfulness might impose a penalty on the quality of the solutions output by deterministic mechanisms. Randomization provably helps to improve the approximation quality, although in order to impose truthfulness we still have to content ourselves with suboptimal allocations. This makes an interesting parallel with the case where the social cost is minimized, as in the latter case optimality and truthfulness can be both enforced by resorting to randomization.

Naturally, our results leave a gap between upper and lower bounds for both deterministic and randomized truthful mechanisms. To close these gaps a better understanding of truthfulness without money of multi-dimensional agents needs to be acquired.

References

Alon, N.; Feldman, M.; Procaccia, A. D.; and Tennenholtz, M. 2010. Strategyproof approximation of the minimax on networks. *Mathematics of Operations Research* 3:513–526.

Dokow, E.; Feldman, M.; Meir, R.; and Nehama, I. 2012. Mechanism design on discrete lines and cycles. In *EC*, 423–440.

Fotakis, D., and Tzamos, C. 2013. On the power of deterministic mechanisms for facility location games. In *ICALP*, 449–460.

Fotakis, D.; Krysta, P.; and Ventre, C. 2014. Combinatorial auctions without money. AAMAS '14, 1029–1036.

Koutsoupias, E. 2014. Scheduling without payments. *Theory Comput. Syst.* 54(3):375–387.

Lu, P.; Wang, Y.; and Zhou, Y. 2009. Tighter bounds for facility games. In *WINE*, 137–148.

Moulin, H. 1980. On strategy-proofness and single-peakedness. *Public Choice* 35:437–455.

Nisan, N., and Ronen, A. 2001. Algorithmic Mechanism Design. *Games and Economic Behavior* 35:166–196.

Procaccia, A. D., and Tennenholtz, M. 2013. Approximate mechanism design without money. *ACM Trans. Economics and Comput.* 1(4):18.

Schummer, J., and Vohra, R. V. 2002. Strategy-proof location on a network. *Journal of Economic Theory* 104:405–428.

Serafino, P., and Ventre, C. 2014. Heterogeneous facility location without money on the line. In *ECAI*, 807–812.