ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-807

Heterogeneous Facility Location without Money on the Line

Paolo Serafino[1] and **Carmine Ventre**[1]

Abstract. The study of facility location in the presence of self-interested agents has recently emerged as the benchmark problem in the research on mechanism design without money. Here we study the related problem of *heterogeneous 2-facility location*, that features more realistic assumptions such as: (*i*) multiple heterogeneous facilities have to be located, (*ii*) agents' locations are common knowledge and (*iii*) agents bid for the set of facilities they are interested in. We study the approximation ratio of both deterministic and randomized truthful algorithms when the underlying network is a line. We devise an $(n-1)$-approximate deterministic truthful mechanism and prove a constant approximation lower bound. Furthermore, we devise an optimal and truthful (in expectation) randomized algorithm.

1 Introduction

Mechanism design without money is a relatively recent and challenging research agenda introduced by Procaccia and Tennenholtz in [8]. It is mainly concerned with the design of *truthful*[2] (or *strategyproof*, *SP* for short) *mechanisms* in scenarios where monetary compensation cannot be used as a means to realign the agents' interest to the mechanism designer's objective (as, e.g., done by VCG mechanisms). It has been noticed that such a circumstance occurs very frequently in real-life scenarios, as payments between agents and the mechanism are either illegal (e.g., organ transplant) or unethical (e.g., in the case of political decision making). To circumvent the impossibility of utilizing payments to enforce truthfulness, Procaccia and Tennenholtz propose instead to leverage the *approximation ratio* of the mechanism in those cases where the optimal outcome is not truthful. The facility location problem is arguably the archetypal problem in mechanism design without money [8]. It demands locating a set of facilities on a network, on input the bids of the agents for their locations, in such a way as to minimize the total connection cost of the agents (i.e., the sum of the distances of each agent to the *nearest facility*). If we regard the problem of locating facilities as a political decision (e.g., a city council locating facilities of public interest on the basis of the population residing in a certain area), the impossibility to utilize payments in this context becomes immediately apparent.

Inspired by the facility location problem, and aiming at analyzing a richer and more realistic setting, we introduce and study the *heterogeneous 2-facility location without money*.[3] In detail, it demands locating 2 *heterogeneous facilities* (i.e., serving different purposes) on a network on input the bids of the agents for the facilities they are interested in, the aim being that of *minimizing the connection cost of the agents to the facilities they bid for*. We study the approximation ratio of truthful deterministic mechanisms when agents are located on a line, prove that *the optimal algorithm is not truthful* (by giving a lower bound of $9/8$) and propose a *truthful* $(n-1)$*-approximate deterministic algorithm*. In order to provide better approximation guarantees we then turn our attention to randomized algorithms. We devise an *optimal randomized algorithm* and prove it is *truthful in expectation*.

The remainder of this paper is organized as follows. §2 is devoted to survey some related literature. In §3 we formalize our model for the heterogeneous facility location problem on the line. In §4 we discuss our results about deterministic algorithms whereas in §5 we present our results for randomized algorithms.

2 Related Work

The facility location problem has proved a fertile research problem and, as such, has been addressed by various research communities.

The Social Choice community has been mostly concerned with the problem of locating a single facility on the line. In his classical paper [7] Moulin characterizes the class of generalized median voter schemes as the only deterministic *SP* mechanisms for *single-peaked* agents on the line. Schummer and Vohra [9] extend the result of Moulin to the more general setting where *continuous graphs* are considered, characterizing *SP* mechanisms on *continuous lines and trees*. They show that on circular graphs every *SP* mechanism must be dictatorial.

From a Mechanism Design perspective, the aforementioned paper [8] initiated the field of approximate mechanism design without money. For the 2-facility location problem, 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 $(n-2)$-approximate, where n is the number of agents. Furthermore, they provide a lower bound of $3/2$ on the approximation ratio of any *SP* algorithm for the facility location problem on the line and conjecture a lower bound of $\Omega(n)$. The latter conjecture has been recently proven by Fotakis et al. [3]. Their main result is the characterization of deterministic *SP* mechanisms with *bounded approximation ratio* for the 2-facility location problem on the line. They show that there exist only two such algorithms: (*i*) a mechanism that admits a unique dictator or (*ii*) the Two-Extremes mechanism proposed in [8]. Lu et al. [6], improve several bounds studied in [8]. Particularly, they prove a 1.045 lower bound for randomized mechanisms for the 2-facility location problem on the line, and present a randomized $n/2$-approximate mechanism. Alon et al. [1] derive a linear (in the number of agents) lower bound for *SP* mechanisms on continuous cycles. Furthermore, they derive a constant approximation bound for randomized mechanisms in the same settings. Dokow

[1] SCM, Teesside University UK, email: {p.serafino, c.ventre}@tees.ac.uk
[2] A mechanism is *SP* if truthtelling is a dominant strategy for agents. See §3.
[3] The present research agenda and some preliminary results were sketched in the extended abstract [10].

et al [2] shift the focus of research to *discrete* lines and cycles instead. They prove that *SP* mechanisms on *discrete large cycles* are nearly-dictatorial in that all agents can effect the outcome to a certain extent. Contrarily to the case of continuous cycles studied in [9], for small discrete graphs Dokow et al. prove that there are anonymous *SP* mechanisms. Furthermore, they prove a linear lower bound in the number of agents for the approximation ratio of *SP* mechanisms on discrete cycles. Another interesting line of research in this area advocates the use of *imposing mechanisms*, i.e. mechanisms able to limit the way agents exploit the outcome of a game. For the facility location problem, imposing mechanisms typically prevent an agent from connecting to some of the facilities, thus increasing her connecting cost and penalizing liars. Following this wake, in [4] Fotakis et al. consider *winner-imposing* mechanisms, namely mechanisms that (*i*) allocate a facility only at a location where there is an agent requesting it (as opposed to mechanisms that allocate facilities at arbitrary locations) and (*ii*) require that an agent that *wins* a facility (i.e. has a facility allocated to her location) must connect to it. Fotakis et al. prove that the winner-imposing version of the Proportional Mechanism proposed in [5] is *SP* for the K-facility location problem and achieves an approximation ratio of at most $4K$, for $K \geq 1$. Furthermore they propose a deterministic non-imposing group strategyproof $\mathcal{O}(\log n)$-approximate mechanism for a variant of the facility location problem on the line with opening costs of facilities and no constraint on the number of facilities to be located.

3 Model and Preliminary Definitions

The *heterogeneous 2-facility location problem on the line* (hereinafter facility location, for short) consists of locating facilities on a *linear unweighted graph*. More specifically, we are given a set of agents $N = \{1, \ldots, n\}$; an undirected unweighted linear graph $G = (V, E)$, where $V \supseteq N$; a set of facilities $\mathfrak{F} = \{\mathsf{F}_1, \mathsf{F}_2\}$.

Agents' *types* are subsets of $\mathfrak{F}$, called their *facility set*. We denote the true type of agent i as $T_i \subseteq \mathfrak{F}$.[4] 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 $F_i \in V$ 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. Naturally, agents seek to minimize their cost. Therefore, they could 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.

A mechanism M is *truthful* (or *strategyproof*, *SP*, for short) if for any i, declarations of the other agents, denoted as $\mathcal{T}_{-i}$, and 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 a *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 minimize the *social cost* function $cost(\mathcal{F}) = \sum_{i=1}^{n} cost_i(\mathcal{F})$, namely: $M(\mathcal{T}) \in \operatorname{argmin}_{\mathcal{F} \text{ feasible}} cost(\mathcal{F})$. We call these mechanisms *optimal* and denote an optimal allocation on declaration vector $\mathcal{T}$ as $OPT(\mathcal{T})$ if $cost(OPT(\mathcal{T})) = \min_{\mathcal{F} \text{ feasible}} cost(\mathcal{F})$. Alas, sometimes we have to content ourselves with sub-optimal solutions. In particular, we say that a mechanism M is α-approximate if $cost(M(\mathcal{T})) \leq \alpha \cdot cost(OPT(\mathcal{T}))$. 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').

[4] Sometimes, slightly abusing notation, we will regard T_i as a set of *indices* j s.t. $\mathsf{F}_j \in T_i$.

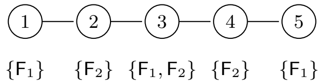

Figure 1. Instance showing that OPT is not truthful

4 Deterministic Mechanisms

In this section we study deterministic mechanisms for the 2-facility location problem. We first ask ourselves whether the optimal allocation for the facility location problem is truthful, to which we give a negative answer in Theorem 1 and provide also a lower bound of 9/8 for the approximation of deterministic SP algorithms. Afterwards, we discuss an $(n-1)$-approximate deterministic algorithm for the facility location problem.

Theorem 1. No deterministic α-approximate SP mechanism can obtain an approximation ratio $\alpha < 9/8$.

Proof. Let us consider the instance depicted in Figure 1 according to the following declarations: $T_1 = \{\mathsf{F}_1\}$, $T_2 = \{\mathsf{F}_2\}$, $T_3 = \{\mathsf{F}_1, \mathsf{F}_2\}$, $T_4 = \{\mathsf{F}_2\}$, $T_5 = \{\mathsf{F}_1\}$. It can be easily checked that the optimal locations for this instance are the ones that locate a facility on node 3 and the other on either node 2 or 4, namely: $(F_1^* = 2, F_2^* = 3)$, $(F_1^* = 4, F_2^* = 3)$, $(F_1^* = 3, F_2^* = 2)$ and $(F_1^* = 3, F_2^* = 4)$. Let us note that any α-approximate algorithm with $\alpha < 9/8$ on input $\mathcal{T}$ would return an optimal solution. Indeed, it can be easily checked that the two second-best solutions $(F_1 = 2, F_2 = 4)$ and $(F_1 = 4, F_2 = 2)$ are $8/7$-approximate, being their cost 8 whereas $cost(OPT(\mathcal{T})) = 7$.

Let us consider the optimal solution $(F_1^* = 2, F_2^* = 3)$. If agent 5 reports $T_5' = \{\mathsf{F}_1, \mathsf{F}_2\}$, then the only optimal solution is $OPT(T_i', \mathcal{T}_{-i}) = (3, 4)$. We note that, since the cost (with respect to $(T_i', \mathcal{T}_{-i})$) of this optimal solution is 8 whereas the cost of any second best solution (i.e, $(F_1 = 4, F_2 = 3)$, $(F_1 = 2, F_2 = 3)$ and $(F_1 = 2, F_2 = 4)$) is 9, any α-approximate algorithm with $\alpha < 9/8$ would return the optimum. Furthermore, we note that the optimal solution is not *SP*, since $cost_5(OPT(T_i', \mathcal{T}_{-i})) = 2 < 3 = cost_5(OPT(\mathcal{T}))$. We note that, due to the intrinsic symmetry of the instance, a similar argument applies for solution $(F_1^* = 4, F_2^* = 3)$ when agent 1 reports $T_1' = \{\mathsf{F}_1, \mathsf{F}_2\}$.

Let us consider the optimal solution $(F_1^* = 3, F_2^* = 4)$. If agent 2 reports $T_2' = \{\mathsf{F}_1, \mathsf{F}_2\}$, then the only optimal solution is $OPT(T_i', \mathcal{T}_{-i}) = (F_1^* = 2, F_2^* = 3)$. We note that, since the cost (with respect to $(T_i', \mathcal{T}_{-i})$) of this optimal solution is 7 and the cost of any second best solution (i.e, $(F_1 = 2, F_2 = 4)$, $(F_1 = 3, F_2 = 4)$ and $(F_1 = 4, F_2 = 3)$) is 8, any α-approximate algorithm with $\alpha < 9/8$ would return the optimum. Furthermore, we note that the optimal solution is not *SP*, since $cost_2(OPT(T_i', \mathcal{T}_{-i})) = 1 < 2 = cost_2(OPT(\mathcal{T}))$. We note that, due to the intrinsic symmetry of the instance, a similar argument applies for solution $(F_1^* = 3, F_2^* = 2)$ when agent 4 reports $T_4' = \{\mathsf{F}_1, \mathsf{F}_2\}$. $\square$

We now discuss TWOEXTREMES, a deterministic mechanism which is truthful and returns linear-approximate allocations. The algorithm, reported in Algorithm 1, is inspired by Two-Extremes of

[8], the difference being that, due to the multi-dimensional nature of our problem, we need to check for the feasibility of solutions putting facilities at the extremes and handle cases of clash.

Algorithm 1: TWOEXTREMES

Require: Line G, facilities $\mathfrak{F} = \{\mathsf{F}_1, \mathsf{F}_2\}$, declarations $\mathcal{T} = \{T_1, \ldots, T_n\}$
Ensure: $F(\mathcal{T})$, a $(n-1)$-approximate allocation for 2-facility location on G

```
F_1 := min V_1[T]
F_2 := max V_2[T]
if F_1 = F_2 then
    if F_2 − 1 ≠ NIL then
        F_2 := F_2 − 1
    else
        F_1 := F_1 + 1
    end if
end if
return (F_1, F_2)
```

We begin by proving the truthfulness of the algorithm.

Theorem 2. 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 in 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. $\square$

In order to prove the approximation guarantee of TWOEXTREMES, we initially prove a lower bound on the value of the optimal social cost.

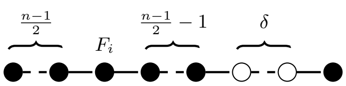

Figure 2. Bounding OPT_i from below

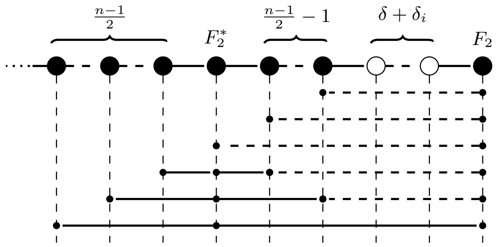

Figure 3. Computing $cost(\mathcal{LR}_2(\mathcal{T}))$. Full edges denote links used by OPT_2 while dashed edges denote links used in $cost(\mathcal{LR}_2(\mathcal{T})) - OPT_2$.

Lemma 3. Let $\mathcal{T}$ be an instance of the 2-facility location problem, such that $n_1 = |V_1|$, $n_2 = |V_2|$, and $\delta > 0$ is the number of empty nodes in between V_1 and V_2. Then the following holds: $cost(OPT(\mathcal{T})) \geq \frac{n_1^2}{4} + \frac{n_2^2}{4} - \frac{1}{2} + 2\delta$.

Proof. Let us take into consideration the minimum-cost instance depicted in Figure 2. In this instance, all but one agents requesting a facility are in a contiguous chain, whereas the isolated agent is at distance δ from the nearest agent (in Figure 2 this pattern is shown for agents in V_i). It can be easily checked that the following holds:

$$OPT \geq 2 \sum_{i=1}^{\frac{n_1-1}{2}} i + \delta + 2 \sum_{i=1}^{\frac{n_2-1}{2}} i + \delta = \frac{n_1^2}{4} + \frac{n_2^2}{4} - \frac{1}{2} + 2\delta,$$

which concludes the proof. $\square$

Theorem 4. Algorithm TWOEXTREMES is $(n-1)$-approximate.

Proof. Let us consider a generic instance $\mathcal{T}$. Moreover, let (F_1^*, F_2^*) be an optimal solution for such an instance, and let $cost(OPT(\mathcal{T})) = OPT_1 + OPT_2$, where $OPT_1 = \sum_{i \in V_1} d(i, F_1^*)$ and $OPT_2 = \sum_{i \in V_2} d(i, F_2^*)$ denote the cost incurred by the agents to connect to facility F_1 and F_2, respectively. Let $\mathcal{LR}(\mathcal{T})$ be the solution output by TWOEXTREMES on input $\mathcal{T}$ and let $(F_1 = \mathcal{LR}_1(\mathcal{T}), F_2 = \mathcal{LR}_2(\mathcal{T}))$ denote the locations that $\mathcal{LR}(\mathcal{T})$ computes for the two facilities. We can express the cost of location (F_1, F_2) as function of the optimal allocation (F_1^*, F_2^*) as follows:

$$\begin{aligned} cost(\mathcal{LR}(\mathcal{T})) = OPT_1 + 2 \sum_{i \in N_1^R \setminus F_1} d(i, F_1) + d(F_1, F_1^*) \\ + OPT_2 + 2 \sum_{i \in N_2^L \setminus F_2} d(i, F_2) + d(F_2, F_2^*), \end{aligned}$$

where N_j^R (N_j^L, respectively) denotes the set of nodes in $V_j[\mathcal{T}]$ to the right (left, respectively) of the median. Figure 3 gives the geometric intuition behind this equality.

We can then observe:

$$\begin{aligned} cost(\mathcal{LR}(\mathcal{T})) \leq & OPT + (n_1 - 3) \cdot d(F_1, F_1^*) + d(F_1, F_1^*) \\ & + (n_2 - 3) \cdot d(F_2, F_2^*) + d(F_2, F_2^*) \\ \leq & OPT + (n-3) \cdot (d(F_1, F_1^*) + d(F_2, F_2^*)) \end{aligned}$$

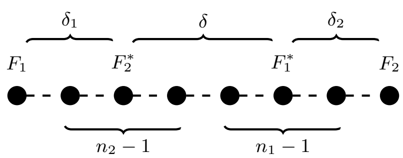

Figure 4. Upper bound to $d(F_1, F_1^*)$ and $d(F_2, F_2^*)$

where: (i) the first inequality follows from upper-bounding $d(i, F_1)$ and $d(i, F_2)$, respectively, by $d(F_1, F_1^*)$ and $d(F_2, F_2^*)$, whereas the second inequality follows from upper-bounding n_1 and n_2 by $n-1$ (i.e., $\max\{n_1, n_2\} \leq n$ since $|V_1| > 0$ and $|V_2| > 0$). In order to upper bound $d(F_1, F_1^*)$ and $d(F_2, F_2^*)$, let us consider the generic instance depicted in Figure 4, where δ is the number of "empty" nodes between F_2^* and F_1^*, δ_1 is the number of empty nodes between F_1 and F_2^* and δ_2 is the number of empty nodes between F_1^* and F_2. It is easy to check that $d(F_1^*, F_1) \leq (\frac{n_1}{2} + n_2 - 1 + \delta + \delta_1)$ and $d(F_2^*, F_2) \leq (\frac{n_2}{2} + n_1 - 1 + \delta + \delta_2)$, which applied to the last inequality yields:

$$cost(\mathcal{LR}(\mathcal{T})) \leq OPT + (n-2)\left(\frac{3}{2}(n_1+n_2) + 2\delta + \delta_1 + \delta_2\right).$$

In virtue of Lemma 3, $\frac{3}{2}(n_1 + n_2) + 2\delta + \delta_1 + \delta_2$ is bounded from above by OPT. Applying the above lower bound to the last inequality yields the following: $cost(\mathcal{LR}(\mathcal{T})) \leq (n-1) \cdot OPT$ which proves the claim. □

We finish this section by proving that the analysis of TwoExtremes presented above is tight.

Theorem 5. The upper-bound for the TwoExtremes algorithm is tight.

Proof. We are going to exhibit an instance for which the TwoExtremes algorithm obtains an approximation ratio of $(n-1)$. The instance we consider is the one depicted in Figure 5 and is such that $|V_1| = n$, $|V_2| = 1$ and n is odd. The number of nodes of the graph is $n + \delta$, where δ is the number of empty nodes. The declarations, depicted in brackets below each nodes are as follows: $T_i = \{\mathsf{F}_1\}$ for each $1 \leq i < n$, $T_n = \{\mathsf{F}_1, \mathsf{F}_2\}$. As before, (F_1^*, F_2^*) and (F_1, F_2) denote the optimal allocation and the outcome of the TwoExtremes algorithm, respectively. It is easy to check that (1) gives the cost of the optimal location, whereas (2) gives the cost of (F_1, F_2):

$$cost(OPT(\mathcal{T})) = 2 \cdot \sum_{i=1}^{\frac{(n-1)}{2}} (i) + \delta = \frac{n^2 - 1 + 4\delta}{4} \tag{1}$$

$$\begin{aligned} cost(\mathcal{LR}(\mathcal{T})) &= \sum_{i=1}^{n-1} (\delta + i + 1) \\ &= \frac{n^2 - 3n + 2(n-1)\delta - 2}{2}. \end{aligned} \tag{2}$$

Equation (3) below expresses the approximation ratio of the TwoExtremes algorithm with respect to the instance of Figure 5 as a function of both the number of players n and the number of empty nodes δ.

$$\alpha(n, \delta) = 2 \cdot \frac{n^2 - 3n + 2(n-1)\delta - 2}{n^2 - 1 + 4\delta} \tag{3}$$

We can see from (3) that if $\delta \in \omega(n^2)$ then $\alpha(n, \delta)$ tends to $n - 1$. □

It is not hard to check from the analysis above that if there are no empty nodes in the instance, then TwoExtreme returns a *constant* (i.e., 2 as n tends to infinity) approximation of the minimum social cost. This implies that for instances that are not sparse in requests, the gap between our bounds becomes slim.

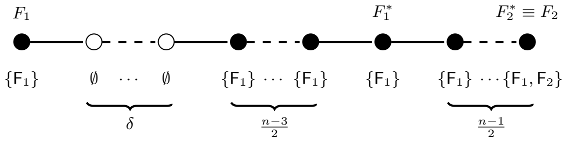

Figure 5. TwoExtremes is $\Theta(n-1)$-approximate

5 Randomized Mechanisms

In this section we present our main result, a truthful randomized optimal algorithm for the 2-facility location problem. The main idea of the algorithm is to *use randomization between optimal outcomes whenever possible*, and to adopt a *truthful-preserving allocation policy* whenever the set of optimal solutions is too small to allow randomization. To describe the algorithm, it is important to define a couple of concepts of interest.

Let S_k be the set of *optimal locations* taking into consideration the requests for facility F_k alone.[5] By the results in [7], we know that the optimal location for a single facility is a median and therefore the set of the optimal locations is either a singleton, i.e. when the number of requests is an odd number, or has size greater than 1, i.e. when the number of requests is even.

A solution F_k is *extreme* for S_k w.r.t. S_{k+1} if: (i) $|S_k| = 2$, (ii) $|S_k \cap S_{k+1}| \leq 1$ and (iii) $F_k = \text{argmax}_{\ell \in S_k}\{d(\ell, S_{k+1})\}$, where $d(\ell, S_{k+1}) = \min_{s \in S_{k+1}} d(\ell, s)$.

Let O denote the set of optimal allocations for the 2-facility location problem. $\mathcal{M} \subseteq O$ is a *mean set* for S_k if the *expected value* of F_k when a solution is drawn uniformly at random from $\mathcal{M}$ equals the average over S_j. More formally, $\mathcal{M} \subseteq O$ is a mean set for S_k if $\underset{U}{E}[F_k] = \text{avg}(S_k) = \frac{\max(S_k)+\min(S_k)}{2}$, where U is the uniform distribution defined over $\mathcal{M}$. Furthermore, an allocation for F_k that is drawn uniformly at random from a mean set $\mathcal{M}$ for S_k will be referred to as *mean solution* for S_k. For the sake of notational conciseness, when referring to extreme and mean solutions we omit S_k and S_{k+1} as they can be easily deduced from the context.

Lemma 6. If $i \in V_k$ and $V_k' = V_k \setminus \{i\}$, then $d(i, \mathit{avg}(S_k)) < d(i, \mathit{avg}(S_k'))$.

Proof. Let us consider the case when $i \leq \min\{S_k\}$, the case when $i \geq \max\{S_k\}$ is symmetric. If $|S_k| = 1$, let s_k denote the sole element of S_k. If $i \notin V_k'$, then $|S_k'| > 1$ and S_k' is such that $\min(S_k') = s_k$ and $\max(S_k') = \ell$, where $\ell \in V_k'$ is the location of the leftmost agent such that $\ell > s_k$ and $k \in T_\ell$. Clearly, $i \leq \text{avg}(S_k) < \text{avg}(S_k')$, which implies the claim. If $|S_k| > 2$, then $|S_k'| = 1$. If $i \notin V_k'$ then $S_k' = \{\max(S_k)\}$ from which it follows that $i \leq \text{avg}(S_k) < \text{avg}(S_k')$, and the claim. □

Lemma 7. Let $i \in N$ be an agent such that $i \in V_k$, $V_k' = V_k \setminus \{i\}$. Then $\min_{\ell \in S_k}\{d(i, \ell)\} \leq \min_{\ell' \in S_k'}\{d(i, \ell')\}$.

Proof. Let us assume that $i \leq \min(S_k)$. S_k can be either a singleton (if $|V_k|$ is odd) or have cardinality greater than 1 (if $|V_k|$ is even). Let $S_k = \{s_k\}$, then $|S_k'| > 1$. Let $r = \max(S_k')$. The thesis holds since $\min_{\ell \in S_k} |i - \ell| = |i - s_k| = \min_{\ell' \in S_k'} |i - \ell'|$. If $|S_k| > 1$, let $l = \min(S_k)$ and $r = \max(S_k)$. Then $S_k' = \{r\}$. The thesis holds in this case since $\min_{\ell \in S_k} |i - \ell| = |i - l| < |i - r| = \min_{\ell' \in S_k'} |i - \ell'|$. The same argument holds for the case when $i \geq \max(S_k)$. Finally, we observe that when $\min(S_k) < i < \max(S_k)$ then $i \notin V_k$. □

[5] For notational convenience, in this section we let the index of the two facilities be binary and all the operations involving indexes be modulo 2. Hence, we will refer indistinctly to one facility as F_k and to the other one as F_{k+1}.

In essence, the previous lemma states that in a monodimensional setting if an agent does not declare a facility to which she is interested in, the space of optimal allocation points gets further away from her.

Lemma 8. Let F_k and F'_k be two extreme solutions. Then it must be $F_k = F'_k$.

Proof. We note that since F_k and F'_k are by hypothesis two extreme solutions it must be that $S_k = S'_k = \{l, r\}$. Let us suppose w.l.o.g. that $F_k = r$. Since both F_k and F'_k are extreme solutions, it must be the case that $|S_k \cap S_{k+1}| \leq 1$ and $|S'_k \cap S'_{k+1}| \leq 1$. This implies that $s \leq l < r$, where s is the element of S_{k+1} nearest to S_k. Let us suppose, for the sake of contradiction, that $F'_k = l$. In this case it must be $l < r \leq s'$, where s' is the element of S'_{k+1} nearest to S_k. We observe that whenever this happens $s \in S'_{k+1}$, which implies that $|S_k \cap S'_{k+1}| \geq 2$ and contradicts the hypothesis that F'_k is an extreme solution. $\square$

The previous lemma essentially states that an agent cannot gain on a facility assigned as an extreme solution, unless she changes the declaration for that facility.

Lemma 9. Let $|S_k| = 1$, F'_k be an extreme solution for S'_k, and let $V'_k = V_k \setminus \{i\}$. Then $d(F_k, i) \leq d(F'_k, i)$.

Proof. Since $F_k = \min_{\ell \in S_k}\{d(i, \ell)\}$, and since, in the best case for agent i, $F'_k = \min_{\ell' \in S'_k}\{d(i, \ell')\}$, by Lemma 7 $d(i, F_k) \leq d(i, F'_k)$. $\square$

We can now discuss the algorithm RANDOPT. Algorithm 2 reports the pseudocode for RANDOPT. Algorithm RANDOPT makes use of procedure COMPUTEMEANSET, which takes as input two locations $\mathcal{L}_k$ and $\mathcal{L}_{k+1}$ and returns a mean set $\mathcal{M}$ such that $\underset{U}{E}[F_k] = \mathcal{L}_k$ and $\underset{U}{E}[F_{k+1}] = \mathcal{L}_{k+1}$. We point out that the proof of Theorem 10 provides a constructive and efficient way of computing mean set $\mathcal{M}$.

Algorithm 2: RANDOPT

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

```
1: ∀k S_k := Opt(V_k[T])
2: if ∃k ∈ {0,1} s.t. |S_k| = 2 and |S_k ∩ S_{k+1}| ≤ 1 then
3:    F_k := argmax_{v∈S_k} { d(v, S_{k+1}) }
4:    if |S_{k+1}| = 2 then
5:       F_{k+1} := argmax_{v∈S_{k+1}} { d(v, S_k) }
6:       return (F_k, F_{k+1}) w.p. 1
7:    else
8:       M := COMPUTEMEANSET(F_k, avg(S_{k+1}))
9:       return (F_k, F_{k+1}) ∈ M w.p. 1/M
10:   end if
11: else
12:   M := COMPUTEMEANSET(avg(S_k), avg(S_{k+1}))
13:   return (F_k, F_{k+1}) ∈ M w.p. 1/M
14: end if
```

We are now going to prove two important properties of algorithm RANDOPT.

Theorem 10. Algorithm RANDOPT always returns an optimal solution.

Proof. It is easy to check that RANDOPT returns either a mean solution or an extreme solution. We are going to prove now that the solutions returned by the algorithm are actually feasible. It is easy to see that solutions returned by RANDOPT are always feasible whenever $S_k \cap S_{k+1} = \emptyset$, so in the remainder we are going to assume that $S_k \cap S_{k+1} \neq \emptyset$. We need to consider three cases: (c.1) both facilities are allocated as extreme solutions (Line 6), denoted in the sequel as (E, E); (c.2) one facility is allocated as an extreme solution while the other facility is allocated as a mean solution (Line 9), referred to as either (E, M) or (M, E); and (c.3) both facilities are allocated as mean solutions (Line 13), denoted as (M, M).

In Line 6 (case c.1) the algorithm allocates both facilities as extreme solutions, so $|S_k| = 2$, $|S_{k+1}| = 2$ and $|S_k \cap S_{k+1}| \leq 1$. Let us suppose w.l.o.g. that $S_k = \{l, l+1\}$ $S_{k+1} = \{l+1, l+2\}$. It is easy to check that $(l, l+2)$, where the first (second, respectively) element of the ordered couple denotes the location of facility F_k (F_{k+1}, respectively), is a feasible extreme solution for F_k and F_{k+1}.

In Line 9 (case c.2) the algorithm allocates a facility as an extreme solution and the other one as a mean solution. W.l.o.g. let us suppose that F_k is allocated as an extreme solution and F_{k+1} is allocated as a mean solution. Therefore, we have $|S_k| = 2$, $|S_k \cap S_{k+1}| \leq 1$ and $|S_{k+1}| \neq 2$. Let us denote $S_k = \{l, l+1\}$ and let us suppose w.l.o.g. that $S_k \cap S_{k+1} = \{l+1\}$ (i.e., the case when $S_k \cap S_{k+1} = \{l\}$ is symmetric). There are two cases to consider: (*i*) $|S_{k+1}| = 1$, (*ii*) $|S_{k+1}| > 2$. We notice that in both cases $F_k = l$ is a feasible extreme solution for S_k. When case (*i*) occurs, then $S_{k+1} = \{l+1\}$ and $\mathcal{M} = \{(l, l+1)\}$ is a feasible mean set for S_{k+1}. When case (*ii*) occurs, $\mathcal{M} = \{(l, \min(S_{k+1})), (l, \max(S_{k+1}))\}$ is a feasible mean set for S_{k+1}.

In Line 13 (case c.3) the algorithm returns an (M, M) solution, so either (*i*) $|S_k| \neq 2$ and $|S_{k+1}| \neq 2$ or (*ii*) $|S_k \cap S_{k+1}| > 1$. Let us consider case (*i*). Let us suppose that $|S_k| > 2$. If allocations $(\min(S_k), \min(S_{k+1}))$ and $(\max(S_k), \max(S_{k+1}))$ are in O then $\mathcal{M} = \{(\min(S_k), \min(S_{k+1})), (\max(S_k), \max(S_{k+1}))\}$ is trivially a mean set, and the claim is true. The same holds if $(\min(S_k), \max(S_{k+1}))$ and $(\max(S_k), \min(S_{k+1}))$ are in O. If neither of the previous holds, then $\min(S_{k+1}) = \max(S_{k+1})$, hence $S_{k+1} = \{s\}$ and $s \in \{\min(S_k), \max(S_k)\}$. Then both $(\min(S_k)+1, s)$ and $(\max(S_k)-1, s)$ are in O and $\mathcal{M} = \{(\min(S_k)+1, s), (\max(S_k)-1, s)\}$ is a mean set. Let us consider the case when $|S_k| = 1$, and let $S_k = \{s\}$. If $|S_{k+1}| = 1$ then $S_k = S_{k+1}$. We note that in this case $\{(s, s-1), (s, s+1)\}$ is a feasible mean set for both S_k and S_{k+1}. If $|S_k| = 1$ and $|S_{k+1}| > 2$, this case is analogous to the case when $|S_k| > 2$ and $|S_{k+1}| = 1$ that we analysed above. Let us now consider case (*ii*). Since $|S_k \cap S_{k+1}| > 1$, then $|S_k| \geq 2$ and $|S_{k+1}| \geq 2$. Then, either $\{(\min(S_k), \min(S_{k+1})), (\max(S_k), \max(S_{k+1}))\}$ or $\{(\min(S_k), \max(S_{k+1})), (\max(S_k), \min(S_{k+1}))\}$ is a feasible mean set for both S_k and S_{k+1}. $\square$

We now prove that the algorithm is truthful.

Theorem 11. Algorithm RANDOPT is SP.

Proof. Consider the outcomes $\mathcal{F} = $ RANDOPT$(\mathcal{T})$ and $\mathcal{F}' = $ RANDOPT$(T'_i, \mathcal{T}_{-i})$. We next show that $cost_i(\mathcal{F}) \leq cost_i(\mathcal{F}')$. Assume by contradiction that $cost_i(\mathcal{F}) > cost_i(\mathcal{F}')$; this implies that there exists at least a facility $k \in \{0, 1\}$ such that $d(i, E[F_k]) > d(i, E[F'_k])$, where F_k (F'_k, respectively) denotes the position of facility k in $\mathcal{F}$ ($\mathcal{F}'$, respectively). In the remainder we will denote as S_k and S'_k the optimal locations of facility k in the instances $\mathcal{T}$ and $(T'_i, \mathcal{T}_{-i})$, respectively.

We have already noticed that algorithm RANDOPT returns either an extreme solution or a mean solution. We denote a possible *output transition* of RANDOPT as $(F_0, F_1) \to (F'_0, F'_1)$, where the left-hand side pair denotes the outcome of the algorithm when each agent reports truthfully, whereas the right-hand side pair denotes the outcome of the algorithm when agent i misreports her type. It can be easiliy showed that all possible output transitions of algorithm RANDOPT can be represented by the directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, such that $\mathcal{V} = \{(E,E),(E,M),(M,E),(M,M)\}$ and $\mathcal{E} = \mathcal{V} \times \mathcal{V} \setminus \{((E,E),(E,E))\}$, i.e. the set of arcs of G comprises all possible transitions but $(E,E) \to (E,E)$. We are going to prove that transition $(E,E) \to (E,E)$ cannot occur. Firstly, we notice that if a solution of type (E,E) is returned then either $S_k \neq S'_k$ or $S_{k+1} \neq S'_{k+1}$. Let us suppose w.l.o.g. that $S_k \neq S'_k$ (i.e. the case when $S_{k+1} \neq S'_{k+1}$ is symmetric). If F_k is an extreme solution for S_k and $S_k \neq S'_k$ then F'_k is a mean solution for S'_k, which would result in a transition $(E,E) \to (M,E)$. To prove the claim, we are now going to prove that every arc of G represents an *SP* transition.

It can be easily verified that transition $(M,M) \to (M,M)$ is *SP* by Lemma 6, whereas transitions $(M,E) \to (M,E)$ and $(E,M) \to (E,M)$ are *SP* by Lemmata 6 and 8.

We note that we can regard $(M,M) \to (E,M)$ and $(M,M) \to (M,E)$ as one case, as in both cases one facility makes a transition $M \to M$ and the other one makes a transition $M \to E$. Lemma 6 assures that transition $M \to M$ is *SP*. Let us focus then on transition $M \to E$. Two cases can occur: (*i*) $S_k = S'_k$, in which case it must be $|S_k| = |S'_k| = 2$ and (*ii*) $S_k \neq S'_k$, in which case $|S_k| = 1$ and $|S'_k| = 2$. In case (*i*), let $S_k = \{l, l+1\}$. We notice that $|S_k \cap S_{k+1}| > 1$ and $E[F_k] = l + \frac{1}{2}$ must hold. Let us suppose w.l.o.g. $i \leq l$ (i.e. the case when $i \geq l+1$ is symmetric). We note that i can gain on k only if $F'_k = l$, which implies that $l < l+1 \leq s$ where s is the nearest point of S'_{k+1} to S_k. This can only happen if $\mathsf{F}_{k+1} \in T_i$ and $\mathsf{F}_{k+1} \notin T'_i$. It follows that $E[|F'_k - i|] = E[|F_k - i|] - \frac{1}{2}$ but $E[|F'_{k+1} - i|] \geq E[|F_{k+1} - i|] + \frac{1}{2}$, which implies that $d(i, E[F'_k]) + d(i, E[F'_{k+1}]) \geq d(i, E[F_k]) + d(i, E[F_{k+1}])$. In case (*ii*), we note that $\mathsf{F}_k \in T_i$ (i.e., if $\mathsf{F}_k \notin T_i$ the location of facility F_k is irrelevant to agent i) and $\mathsf{F}_k \notin T'_i$. By Lemma 9 this transition is *SP*.

We note that we can regard cases $(E,M) \to (E,E)$ and $(M,E) \to (E,E)$ in the same way, as in both cases we have a transition $E \to E$ and a transition $M \to E$. We notice that transition $E \to E$ is *SP* by Lemma 8. Let us now focus on transition $M \to E$. We notice that in this case $|S_k| = 1$ and F'_k is an extreme solution. By Lemma 9 this transition is *SP*.

We note we can regard $(E,E) \to (E,M)$ and $(E,E) \to (M,E)$ as one case, in both cases we have a transition $E \to E$ and a transition $E \to M$. For the transition $E \to E$, we note that Lemma 8 ensures truthfulness. Let us now analyse transition $E \to M$. Agent i can only gain if $\mathsf{F}_k \in T_i$, so the only possible lie is $T'_i = T_i \setminus \{k\}$. Since F_k is an extreme solution, $S_k = \{l, l+1\}$. Let us suppose w.l.o.g. that $i \leq l < l+1$ (i.e., the case when $l < l+1 \leq i$ is symmetric). It is easy to check that $S'_k = \{l+1\}$ and $i < F_k \leq F'_k$, which implies that $d(i, F_k) \leq d(i, F'_k)$.

Let us now consider the case $(M,M) \to (E,E)$. We notice that in this case $|S_k| \neq 2$ and $|S_{k+1}| \neq 2$. To prove this, let us suppose for the sake of contradiction that $|S_k| = 2$. In order to have an (M,M) pair it must be the case that $|S_k \cap S_{k+1}| > 1$ which implies that $|S_{k+1}| \geq 2$. We notice that in this case $|S'_{k+1}| = 1$, which would not result in a (E,E) pair. Let us then consider the case when $|S_k| \neq 2$ and $|S_{k+1}| \neq 2$. We highlight that, since $|S_k| \neq 2$ and $|S_{k+1}| \neq 2$ but $|S'_k| = |S'_{k+1}| = 2$, it must be that $V_k \neq V'_k$ and $V_{k+1} \neq V'_{k+1}$, from which it follows that $|S_k| = |S_{k+1}| = 1$. Furthermore it must be the case that $\mathsf{F}_k \in T_i$ (i.e. otherwise the location of facility F_k would be irrelevant for the cost of agent i) and $\mathsf{F}_k \notin T'_i$. We can apply to both $M \to E$ transitions Lemma 9 to show that strategyproofness is preserved.

Let us now consider the case $(E,E) \to (M,M)$. We have $|S_k| = 2$, $|S_{k+1}| = 2$. We are going to prove that $S_k \neq S'_k$ and $S_{k+1} \neq S'_{k+1}$. For the sake of contradiction, if $S_k = S'_k$, then $S_{k+1} \neq S_{k+1}$ and $|S'_{k+1}| = 1$. Since F'_k must be a mean solution for S'_k, it must be that $|S'_k \cap S'_{k+1}| \geq 1$, which is a contradiction, since $|S'_{k+1}| = 1$. Furthermore, we can assume that $\mathsf{F}_k \in T_i$ (i.e., otherwise the location of facility F_k would be irrelevant for agent i) and $\mathsf{F}_k \notin T_i$. We observe that $|S'_k| = |S'_{k+1}| = 1$. Since (in the best case for agent i) $F_k = \min_{\ell \in S_k}\{d(i,\ell)\}$ and since $F'_k = \min_{\ell' \in S'_k}\{d(i,\ell')\}$ (as S'_k is a singleton) by Lemma 7 strategyproofness is preserved.

We note we can regard $(E,M) \to (M,E)$ and $(M,E) \to (E,M)$ as the same case, since both cases have a transition $E \to M$ for one facility and a transition $M \to E$ for the other one. To fix ideas, let us assume facility F_k makes transition $E \to M$ and facility F_{k+1} makes transition $M \to E$. We are going to prove that $S_k \neq S'_k$ and $S_{k+1} \neq S'_{k+1}$. We note that $|S_k| = 2$ and $|S'_{k+1}| = 2$. Let us suppose for the sake of contradiction that $S_k = S'_k$. We reach a contradiction since $|S'_k| = |S'_{k+1}| = 2$ can yield either a (M,M) solution (if $|S'_k \cap S'_{k+1}| \geq 2$) or a (E,E) solution (if $|S'_k \cap S'_{k+1}| \leq 2$). Let us suppose now that $S_{k+1} = S'_{k+1}$. As before, we reach a contradiction since $|S_k| = |S_{k+1}| = 2$ can yield either a (M,M) solution or a (E,E) solution. Furthermore, it can be easily checked that $|S'_k| = 1$ and $|S_{k+1}| = 1$. We can now analyse each transition singularly. Let us focus on transition $E \to M$. We can restrict ourselves to the case when $\mathsf{F}_k \in T_i$ (i.e., otherwise the location of facility F_k does not affect the cost of agent i) and $\mathsf{F}_k \notin T'_i$. Since (in the best case for agent i) $F_k = \min_{\ell \in S_k}\{d(i,\ell)\}$ and $F'_k = \min_{\ell' \in S'_k}\{d(i,\ell')\}$ (as S'_k is a singleton), strategyproofness is guaranteed by Lemma 7. Let us consider transition $M \to E$. Once again, we can restrict to the case when $\mathsf{F}_{k+1} \in T_i$ and $\mathsf{F}_{k+1} \notin T'_i$. Since $|S_{k+1}| = 1$ and F_{k+1} is an extreme solution, by Lemma 9 strategyproofness is preserved. □

REFERENCES

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

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

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

[4] D. Fotakis and C. Tzamos. Winner-imposing strategyproof mechanisms for multiple Facility Location games. *TCS*, 472:90–103, 2013.

[5] P. Lu, X. Sun, Y. Wang, and Z. A. Zhu. Asymptotically optimal strategy-proof mechanisms for two-facility games. In *EC*, pages 315–324, 2010.

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

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

[8] A. Procaccia and M. Tennenholtz. Approximate mechanism design without money. In *EC*, pages 177–186, 2009.

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

[10] P. Serafino and C. Ventre. Truthful mechanisms for the location of different facilities (Extended Abstract). In *AAMAS*, pages 1613–1614, 2014.