Proceedings Article

Strategyproof Facility Location and the Least Squares Objective

MICHAL FELDMAN, Hebrew University of Jerusalem and Harvard University
YOAV WILF, Hebrew University of Jerusalem

We consider the problem of locating a public facility on a tree, where a set of n strategic agents report their *locations* and a mechanism determines, either deterministically or randomly, the location of the facility. The contribution of this paper is twofold.

First, we introduce, for the first time, a general and clean family of strategyproof (SP) mechanisms for facility location on tree networks. Quite miraculously, all of the deterministic and randomized SP mechanisms that have been previously proposed can be cast as special cases of this family. Thus, the proposed mechanism unifies much of the existing literature on SP facility location problems, and simplifies its analysis.

Second, we demonstrate the strength of the proposed family of mechanisms by proving new bounds on the approximation of the *minimum sum of squares* (miniSOS) objective on line and tree networks. For lines, we devise a randomized mechanism that gives 1.5-approximation, and show, through a subtle analysis, that no other randomized SP mechanism can provide a better approximation. For general trees, we construct a randomized mechanism that gives 1.83-approximation. This result provides a separation between deterministic and randomized mechanisms, as it is complemented by a lower bound of 2 for any deterministic mechanism. We believe that the devised family of mechanisms will prove useful in studying approximation bounds for additional objectives.

Categories and Subject Descriptors: F.2 [**Theory of Computation**]: Analysis of Algorithms and Problem Complexity; J.4 [**Social and behavioral sciences**]: Economics

Additional Key Words and Phrases: approximate mechanism design without money, facility location, least squares

1. INTRODUCTION

In facility location problems, a social planner has to determine the location of a public facility that needs to serve a set of agents. Once the facility is located, each agent incurs some cost that depends on the distance from her ideal location to the chosen location of the facility. This class of problems is realized in many scenarios, including, for example, locating a server in telecommunication networks or locating a library or a fire station in a road network. Facility location problems arise not only in physical settings like the ones just described, but also in more virtual settings where the agents' opinions or preferences can be represented as their locations, and a single outcome has to be chosen. As an example, consider a set of students sitting in a classroom with an air conditioner, where every student has her most preferred temperature, and a single temperature has to be chosen. In all of these examples, one "location" has to be chosen, and every agent would like it to be as near as possible to her most preferred location.

Michal Feldman's work was partially supported by the Israel Science Foundation (grant number 1219/09), by the Leon Recanati Fund of the Jerusalem School of Business Administration, the Google Inter-university center for Electronic Markets and Auctions, and the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 274919

What is the best method to settle the naturally conflicting preferences of the agents? In other words, what is the best way to determine the location of the facility, given the agents' locations? This question, and related ones, have been extensively studied in the literature, from various conceptual and algorithmic perspectives; see, e.g., Marsh and Schilling [1994], the book by Handler and Mirchandani [1979], and the body of literature concerning the performance of a *Condorcet* point[1] [Bandelt 1985; Bandelt and Labbé 1986; Labbe 1985]. The literature considers both *deterministic* mechanisms, in which the facility location is chosen deterministically, and *randomized* mechanisms, which return a probability distribution over locations.

Starting with the work of Moulin [1980], this problem was also studied from a game-theoretic perspective, where the agents are assumed to be strategic, i.e., report their locations in a way that will minimize their individual costs. The game-theoretic perspective of the facility location problem advanced in two main directions, as described below.

Sufficient and necessary conditions for strategyproofness. The first direction seeks to characterize *strategyproof* (SP) mechanisms; i.e., mechanisms that induce truthful reporting as a dominant strategy. This is a crucial attribute since in its absence, strategic agents may misreport their locations. Moulin [1980] and later Schummer and Vohra [2002] provided characterizations of *deterministic* SP mechanisms on line, tree, and cycle networks. While the work of Moulin [1980] concentrated on general single-peaked[2] preferences [Black 1958], Schummer and Vohra [2002] considered the special case in which the cost incurred by an agent is the length of the shortest path from the facility to her location. These results were later extended to various metric spaces (see, e.g., Border and Jordan [1983]). It should be noted that the characterization of SP mechanisms for facility location settings has been studied thus far mainly with respect to *deterministic* mechanisms. An exception is the work by Ehlers et al. [2002], which characterizes randomized SP mechanisms, but is different from our work in two respects. First, their characterization applies only to a line network, while we are interested in the more general case of a tree. Second, they use a different notion of preferences over probability distribution (in particular, preference over probability distributions is defined in terms of first-order stochastic dominance), and, as a result, use a different notion of strategyproofness. This difference has significance effects, for example, on lower bound results for approximation.

Approximation. The second direction, advocated more recently by Procaccia and Tennenholtz [2009], studies the approximation ratio that can be obtained by an SP mechanism with respect to a given objective function. This agenda, often termed "approximate mechanism design without money", has led to extensive work on several domains, including facility location [Procaccia and Tennenholtz 2009; Alon et al. 2009, 2010; Lu et al. 2009, 2010; Fotakis and Tzamos 2010], machine learning [Dekel et al. 2010], and matching [Ashlagi et al. 2010; Dughmi and Ghosh 2010]. Unlike the traditional motivation for approximation, originating from computational hardness, here, approximation is used to achieve strategyproofness.

The approximation ratio of a mechanism is defined with respect to a given objective function, and the standard worst-case notion is being applied. The two objective functions that have been prominently featured in the literature are the *minisum* (i.e., minimizing the sum of agents' costs) and *minimax* (i.e., minimizing the maximum cost of any agent) functions. The approximation bounds of SP mechanisms clearly depend on the specified objective function. For example, while it is well known that the optimal location on a tree with respect to the minisum function can be obtained by an SP mechanism (in particular, by the *median*), it has been shown by Procaccia and Tennenholtz [2009] that no deterministic

[1]A Condorcet point is a location that is preferred to any other location by more than half the agents.

[2]With single-peaked preferences, every agent is associated with an ideal location, considered to be her *peak*, and the closer the facility is to an agent's peak, the most preferred it is.

(respectively, randomized) SP mechanism on a line can achieve a better approximation than 2 (resp., 1.5) with respect to the minimax function.

A special focus was given to the case of facility location problems on networks, where the agents (as well as the facility) are located on various points of an underlying network topology. Within this literature, the *line* and the *tree* networks received significant attention, as natural topologies. The motivation for the study of line topologies arises naturally from any one-dimensional decision that is made by aggregating agents' preferences, such as the air-conditioning example mentioned above. The tree topology is particularly motivated by applications of communication networks, where a tree topology corresponds to hierarchical networks. Indeed, in computer networks, an agent can easily manipulate its perceived network location by generating a false IP address, and therefore SP mechanisms are desired.

1.1. Our contribution

Focusing on tree and line networks, our work advances the literature in the two directions mentioned above, as detailed below.

Sufficient condition for randomized SP mechanisms. The characterization of SP mechanisms for facility location settings has been studied thus far mainly with respect to *deterministic* mechanisms. In this work, we establish a sufficient condition for *randomized* SP mechanisms on any tree network, which essentially provides a family of randomized SP mechanisms, termed "parameterized boomerang" (PB). Quite miraculously, all of the mechanisms that have been devised recently within the facility location domain (and shown to give tight approximation bounds with respect to various objective functions) can be formulated as special cases of the PB family. Thus, the proposed mechanism unifies much of the existing literature on SP approximation for facility location problems and simplifies its analysis. The generality and strength of Mechanism PB is further illustrated through the analysis of the least squares objective (see next paragraph), where various instances of Mechanism PB are shown to achieve good approximation results on line and tree networks. Thus, the sufficient condition provided here is not only interesting in itself but also equips one with a large family of SP mechanisms, and can be served as a useful tool for studying the approximation ratios with respect to various objectives. Mechanism PB is presented in Section 3.

Approximation. In the approximation regime, we study the least squares objective function — minimizing the sum of squares of costs (hereafter *miniSOS*). The miniSOS function is highly relevant in many economic settings, and is related to central notions in other disciplines, such as the *centroid* in geometry, or the *center of mass* in physics. Of particular interest is the applicability of the miniSOS objective to regression learning settings. While it is not immediately obvious, it can be verified that there is a one-to-one mapping between the problems of facility location on a line and regression learning, when restricted to the class of constant functions[3]. Consequently, the prevalence of the miniSOS function in regression learning motivates the study of this objective in our study. In addition, the miniSOS objective has been given a rigorous foundation by Holzman [1990] using an *axiomatic* approach. In particular, Holzman formulated three axioms[4] for locating a facility on a line

[3]Consider a regression learning problem, in which every data point is represented by a point (x_i, y_i) in the plane, and the objective is to find the function that *fits* best the collection of the data points. Suppose one needs to find the best function out of the class of *constant* functions c. In this case, the distance between every point (x_i, y_i) to c is $|y_i - c|$, which is independent of x_i. Therefore, The data points can be essentially thought of as a collection of points $\{y_i\}$ (i.e., the projection of the data points on the y-axis), in which case the problem of finding the best constant function reduces to the problem of facility location on the line (where the line is the y-axis).

[4]The three axioms formulated by Holzman are (roughly speaking): (i) unanimity, stating that if all the agents report the same location, then the reported location should be chosen, (ii) Lipschitz, which is a continuity

or a tree, regarded as sensible requirements, and showed that the unique objective function that satisfies the three axioms is the miniSOS objective. Following the above motivation and the axiomatic foundation laid by Holzman, it is only natural to study the approximation that can be obtained by SP mechanisms with respect to the miniSOS function.

We provide approximation results with respect to the miniSOS objective, for line and tree networks, and for deterministic and randomized mechanisms. Notably, all the mechanisms that are devised in this paper are special cases of Mechanism PB. As in other settings [Procaccia and Tennenholtz 2009; Alon et al. 2010; Lu et al. 2010; Ashlagi et al. 2010], randomized mechanisms are shown to provide better bounds than deterministic ones. Our results are as follows:

Line networks (Section 4). We show that the *median* gives a 2-approximation deterministic SP mechanism, and that no deterministic SP mechanism can achieve a better approximation ratio. For randomized mechanisms, we present an SP mechanism that provides a 1.5-approximation; the mechanism chooses the average location with probability $\frac{1}{2}$ and a random dictator with probability $\frac{1}{2}$. Through a novel technique and subtle analysis, we show that this bound is tight. Interestingly, while the minimax and the miniSOS functions induce different optimal solutions, and different optimal SP mechanisms, they admit the exact same approximation bounds with respect to both deterministic and randomized mechanisms.

Tree networks (Section 5). First, we show that the median mechanism gives a 2-approximation with respect to the miniSOS objective. This result is tight with respect to deterministic mechanisms, following the lower bound established on the line. Our main result in the approximation regime is the construction of a randomized SP mechanism that gives a 1.83-approximation for any tree network. The significance of this result stems from the fact that it establishes a separation between deterministic and randomized mechanisms (recall the lower bound of 2 with respect to deterministic mechanisms, even on the line). The proof of the upper bound requires quite complex analysis and various transformation, and is perhaps the main technical contribution of this paper.

All of the missing proofs are deferred to the full version of the paper.

2. MODEL AND PRELIMINARIES

We use the model of Schummer and Vohra [2002], where the network is represented by a graph G, formalized as follows. The graph is a closed, connected subset of Euclidean space $G \subseteq \mathbb{R}^k$. The graph is composed of a finite number of closed curves of finite length, known as the edges[5]. The extremities of the curves are known as vertices (or nodes). An important class of graphs, which is the focus of this paper, is *tree* graphs — graphs that contain no cycles.

The path between two points $a, b \in G$ is denoted by $path_G(a, b)$. The distance between two points $a, b \in G$, denoted $d_G(a, b)$, is the length of the (unique) path between a and b. We extend the definition of distance between points to distance between a point and a path as follows. Given a point $c \in G$ and a path $path_G(a, b)$, the distance between c and $path(a, b)$, denoted $d_G(path_G(a, b), c)$, is the shortest distance between c and any point on $path(a, b)$; i.e., $d_G(path_G(a, b), c) = min_{l \in path_G(a,b)} d_G(l, c)$. When clear in context, we omit the subscript G.

Let $N = \{1, \ldots, n\}$ be a set of agents. We sometime use $[n]$ to denote the set of agents N. Each agent $i \in N$ has an (ideal) location $x_i \in G$ (agents can be located anywhere on G). The collection $\mathbf{x} = (x_1, \ldots, x_n) \in G^n$ is referred to as the *location profile*.

---

requirement; and (iii) invariance, stating that an agent who moves to a location that is equidistant from the outcome from the same direction will not affect the chosen outcome.

[5]Note that while this model is expanding upon the notion of an interval, it is not analyzing full-dimensional, convex subsets of Euclidean space. Rather, travel is restricted to a road network, where convex combinations of locations are typically not feasible.

A *deterministic mechanism* is a function $f : G^n \to G$ that maps the agents' reported locations to the location of a *facility* (which can be located anywhere on G). If the facility is located at $y \in G$, the cost of agent i is the distance between x_i and y; i.e., $cost(y, x_i) = d(y, x_i)$.

A *randomized mechanism* is a function $f : G^n \to \Delta(G)$, which maps location profiles to probability distributions over G (which randomly designate the facility location). Let $P \in \Delta(G)$ be a probability distribution over G. If $f(x) = P$, then the cost of agent i is the expected distance of the facility location from x_i; i.e., $cost(P, x_i) = E_{y \sim P}[cost(y, x_i)]$. When clear in the context, we write $y \sim f(\mathbf{x})$ for ease of presentation.

A mechanism is called *strategyproof* (SP), or *truthful*, if no agent can benefit from misreporting her location, regardless of the reports of the other agents. Formally, in our scenario, this means that for all $\mathbf{x} \in G^n$, for all $i \in N$, and for all $x'_i \in G$, it holds that $cost(f(\mathbf{x}), x_i) \leq cost(f(x'_i, \mathbf{x}_{-i}), x_i)$, where $\mathbf{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ is the profile of all locations, excluding agent i's location.

We next consider deviations by coalitions (i.e., subsets of agents, as opposed to unilateral deviations). A coalitional deviation is said to be *beneficial* if none of the agents in the coalition incurs a higher cost and at least one agent strictly benefits from the deviation. A mechanism is called *group strategyproof* (GSP) if no subset of agents has a beneficial deviation. Formally, this means that for all $\mathbf{x} \in G^n$, for all $S \subseteq N$, and for all $\mathbf{x}'_S \in G^{|S|}$, either $cost(f(\mathbf{x}), x_i) \leq cost(f(\mathbf{x}'_S, \mathbf{x}_{-S}), x_i)$ for every $i \in S$, or there exists an agent $i \in S$ such that $cost(f(\mathbf{x}), x_i) < cost(f(\mathbf{x}'_S, \mathbf{x}_{-S}), x_i)$, where $\mathbf{x}_S$ (respectively, $\mathbf{x}_{-S}$) is the profile of the locations of the agents in S (resp., the agents not in S).

The quality of a facility location is usually evaluated with respect to some target social function. Given a location profile $\mathbf{x} = (x_1, \ldots, x_n)$ and a facility location y, the social cost of y with respect to $\mathbf{x}$ is given by a function $sc(y, \mathbf{x})$. The social cost of a distribution P with respect to $\mathbf{x}$ is $sc(P, \mathbf{x}) = E_{y \sim P}[sc(y, \mathbf{x})]$.

Given a social cost function, location $y \in G$ is said to be *optimal* with respect to a profile $\mathbf{x}$ if $sc(y, \mathbf{x}) = min_{y' \in G} sc(y', \mathbf{x})$. An optimal location is denoted by $Opt(G, \mathbf{x})$. When clear in the context, we simply write Opt. In addition, we often abuse notation and use Opt to refer to the social cost of an optimal location.

A mechanism f is said to provide α-approximation with respect to a social cost function sc if for every graph G and every location profile $\mathbf{x}$, $sc(f(\mathbf{x}), \mathbf{x})/sc(Opt, \mathbf{x}) \leq \alpha$; that is, the mechanism always returns a solution that is an α factor of the optimal solution.

In this paper we are interested in optimizing the sum of squared distances (SOS) function; that is, $sc(y, \mathbf{x}) = \sum_{i \in N} d(y, x_i)^2$. This objective function is extremely important from both normative and positive perspectives, as discussed in the introduction.

Given a profile $\mathbf{x}$, the median of $\mathbf{x}$ in a tree G, denoted by $\mu(G, \mathbf{x})$, is defined as follows. We start from an arbitrary node (induced by G) as a root. Then, as long as the current location has a subtree that contains more than half of the agents, we smoothly move down this subtree. Finally, when we reach a point where it is not possible to move closer to more than half the agents by continuing downwards, we stop and return the current location.

We continue with several graph theoretic definitions and lemmas. At this point, it is necessary to emphasis the difference between a *location profile*, which was defined earlier and is tightly coupled with a set of agents, and a *location vector*, which is a set of locations in the graph.

Definition 2.1. Given a tree G and a point $x \in G$, let $T(G, x)$ be the set of subtrees defined as follows. If x is a tree node (with degree d_x), then $T(G, x) = \{T_1, \ldots, T_{|d_x|}\}$, where T_i is the subtree of descendant i rooted at x. If x is not a node (i.e., it is a point on an edge), then $T(G, x) = \{T_1, T_2\}$, where T_1 and T_2 are the respective left and right subtrees rooted at x.

Definition 2.2. Let G be a tree, $\mathbf{y} \in G^m$ be a location vector, and $\mathbf{w}$ be a probability vector of size m. The weighted average location with respect to $G, \mathbf{y}$ and $\mathbf{w}$, denoted $wAvg(G, \mathbf{y}, \mathbf{w})$, is a point in G which minimizes the weighted sum of squared distances from the locations in $\mathbf{y}$; i.e., $wAvg \in argmin_{l \in G} \sum_{j \in [m]} w_j d(l, y_j)^2$.

The following lemmas will be required in the sequel.

LEMMA 2.3. *Let $\mathbf{y} \in G^m$ be a location vector, and $\mathbf{w}$ a probability vector of size m. It holds that $a = wAvg(G, \mathbf{y}, \mathbf{w})$ if and only if for every $T_j \in T(G, a)$,*

$$\sum_{i \in [m]: y_i \in T_j} w_i d(y_i, a) \leq \sum_{i \in [m]: y_i \notin T_j} w_i d(y_i, a).$$

COROLLARY 2.4. *Let $\mathbf{y} \in G^m$ be a location vector, and $\mathbf{w}$ a probability vector of size m. The weighted average location with respect to $G, \mathbf{y}$ and $\mathbf{w}$ is unique.*

LEMMA 2.5. *Let $\mathbf{y}, \mathbf{y}' \in G^m$ be location vectors, and $\mathbf{w}$ be a probability vector of size m. For every $i \in [m]$, let $\delta_i = d(y_i, y'_i)$. Let $a = wAvg(G, \mathbf{y}, \mathbf{w})$ and $a' = wAvg(G, \mathbf{y}', \mathbf{w})$. It holds that $d(a, a') \leq \sum_{i \in [m]} w_i \delta_i$.*

3. RANDOMIZED SP MECHANISMS ON A TREE

In this section we introduce a family of randomized SP mechanism for locating a facility on a tree. Unless otherwise stated, the graph G in this section is assumed to be a tree.

The following notion of a *boomerang* mechanism[6] is a key concept in our construction.

Definition 3.1. A deterministic mechanism f is said to be a *boomerang* mechanism if for every location profile $\mathbf{x}$, agent i, and point x'_i, $cost(f(\mathbf{x}'), x_i) - cost(f(\mathbf{x}), x_i) = d(f(\mathbf{x}'), f(\mathbf{x}))$, where $\mathbf{x}' = (x'_i, \mathbf{x}_{-i})$.

That is, a boomerang mechanism is one in which a deviating agent fully absorbs the effect of her deviation on the facility location.

Clearly, every boomerang mechanism is SP. Moreover, every boomerang mechanism is GSP, as established by the following proposition.

PROPOSITION 3.2. *Every boomerang mechanism is GSP on a tree.*

PROOF. Let $f(\mathbf{x})$ be a boomerang mechanism, and assume toward contradiction that there exists a coalition of agents $S \subseteq N$, such that deviating from $\mathbf{x}_S$ to $\mathbf{x}'_S$ is beneficial. Let T_S be the subtree induced by the true locations of S, i.e., for each location $\ell \in T_S$, there exist $i, j \in S$ such that $\ell \in path(x_i, x_j)$. If $f(\mathbf{x})$ is located on T_S, then $f(\mathbf{x}'_S, \mathbf{x}_{-S})$ must be more costly with respect to at least one agent in S, contradicting that the deviation is beneficial. Therefore, $f(\mathbf{x})$ must reside outside of T_S. Now consider the agents in S deviating from $\mathbf{x}_S$ to $\mathbf{x}'_S$, one by one. It is easy to verify that by Definition 3.1 in every such iteration, the facility location cannot get closer to T_S. Thus, the final location $f(\mathbf{x}'_S, \mathbf{x}_{-S})$ is at least as far from T_S as $f(\mathbf{x})$, which contradicts the beneficial deviation. □

Several examples of boomerang mechanisms follow: (The proof is left as an exercise to the reader.) (i) *dictatorship*; i.e., where there exists $i \in N$ such that for every $\mathbf{x}$, $f(\mathbf{x}) = x_i$. (ii) *median* (on a tree). (iii) *k'th-location* (on a line); also known as *generalized median* [Moulin 1980].

We are now ready to introduce the family of randomized SP mechanisms for tree networks. This family is presented as a parameterized mechanism, called "parameterized boomerang".

[6] This notion is closely related to the notion of an "uncompromising mechanism" defined in [Border and Jordan 1983].

Mechanism **"parameterized boomerang"** *(PB).* Let $\mathbf{f} = (f_1, \dots, f_m)$ be a collection of boomerang mechanisms. For every $i \in [m]$, let $y_i = f_i(\mathbf{x})$, and let $\mathbf{y} = (y_1, \dots, y_m)$. Let $\mathbf{w}$ be a probability distribution supported on m elements, and let $a = wAvg(G, \mathbf{y}, \mathbf{w})$. The facility location is chosen according to the following probability distribution:

— for every $i \in [m]$, choose $f_i(\mathbf{x})$ with probability $\frac{1}{2}w_i$.
— choose a with probability $\frac{1}{2}$.

We refer to the two components of the probability distribution as the *boomerang* component and *average* component, respectively. Note that every boomerang mechanism is a special case of PB, with $m = 1$.

The following theorem establishes the strategyproofness of Mechanism PB.

THEOREM 3.3. *Mechanism PB is SP.*

PROOF. Assume by way of contradiction that there exists an agent i that can benefit by misreporting her location as x'_i, inducing a location profile $\mathbf{x}' = (x'_i, \mathbf{x}_{-i})$. We quantify the effect of the deviation on the boomerang component and the average component of Mechanism PB. We begin with the boomerang component. For every $j \in [m]$, let $\delta_j = d(f_j(\mathbf{x}'), x_i) - d(f_j(\mathbf{x}), x_i)$ be the additional cost incurred by i due to the deviation, when f_j is chosen. Since f_j is a boomerang mechanism, it holds that $d(f_j(\mathbf{x}'), x_i) - d(f_j(\mathbf{x}), x_i) = d(f_j(\mathbf{x}'), f_j(\mathbf{x})) \geq 0$. Therefore, the additional cost incurred by i due to the boomerang component is $\sum_{j \in [m]} w_j \delta_j = \sum_{j \in [m]} w_j d(f_j(\mathbf{x}'), f_j(\mathbf{x}))$. By Lemma 2.5, the average component reduces agent i's cost by at most $\sum_{j \in [m]} w_j d(f_j(\mathbf{x}'), f_j(\mathbf{x}))$. The assertion of the theorem follows. □

The following assertion is an immediate corollary.

COROLLARY 3.4. *Every fixed probability distribution over PB mechanisms is SP.*

Interestingly, unlike deterministic boomerang mechanisms that are GSP for every tree network, Mechanism PB is not GSP, even on the line. This is established in the following example.

Example 3.5. Let $f(\mathbf{x})$ be the PB mechanism that chooses a random dictator with probability $\frac{1}{2}$, and the average of the reported locations with probability $\frac{1}{2}$. Suppose there are 3 agents that are located at points $0, 1$ and 3 on the line, and consider the deviation in which the agents located at 0 and 1 report the false locations $-\frac{1}{3}$ and $\frac{2}{3}$, respectively. As a result of the false reports, the average component moved from $1\frac{1}{3}$ to $1\frac{1}{9}$. Consequently, the expected cost of the agent located at 0 decreased, while that of the agent located at 1 was not affected. Therefore, this mechanism is not GSP.

In recent years, various SP mechanisms have been proposed in the literature for the facility location problem on the line with the objective of approximating different social objectives, such as the minisum and the minimax functions. The following proposition shows that all of the mechanisms that have been proposed in this context are special cases of Mechanism PB (or a probability distribution over PB mechanisms).

PROPOSITION 3.6. *The following mechanisms on the line are special cases of Mechanism PB (or a probability distribution over PB mechanisms).*

(*1*) k-location. *Examples of this mechanism are the median mechanism, which is known to minimize the sum of distances, and the leftmost agent mechanism, which provides a 2-approximation for the minimax objective [Procaccia and Tennenholtz 2009] (which is tight with respect to deterministic mechanisms).*

(*2*) left-right-middle *(LRM) [Procaccia and Tennenholtz 2009]. LRM chooses the leftmost agent with probability* $\frac{1}{4}$*, the rightmost agent with probability* $\frac{1}{4}$*, and their middle point with probability* $\frac{1}{2}$*. It provides a (tight)* 1.5*-approximation for the minimax objective.*

(*3*) random dictator *(RD). RD chooses every agent with probability* $\frac{1}{n}$*. It provides a* $2-\frac{2}{n}$*-approximation for the minisum objective [Alon et al. 2009]. We will later establish that RD gives a* 2*-approximation for the miniSOS objective (see Theorem 4.5).*

While it is evident from the last proposition that Mechanism PB is very powerful, it imposes a sufficient condition for strategyproofness, but not a necessary one. This is established by the following example.

Example 3.7. Consider the following mechanism on a line. Choose each of the leftmost and rightmost agents with probability $\frac{1}{2n}$, and the center of every two consecutive locations with probability $\frac{1}{n}$. It is easy to verify that this is an SP mechanism, yet it cannot be formulated as a special case of Mechanism PB or a probability distribution over PB mechanisms.

4. SP MECHANISMS ON A LINE

In this section, we study how well SP mechanisms can approximate the miniSOS objective — minimizing the sum of squared distanced — on a line. In the deterministic case, we present a mechanism that provides 2-approximation, and show that no SP deterministic mechanism can achieve a better ratio. In the randomized case, we construct a mechanism that provides 1.5-approximation, and show that this result is tight.

In this section, the graph is essentially the *real* line, $\mathbb{R}$. It is easy to verify that an optimal location in this case is simply the average.

CLAIM 4.1. *Given a location profile* $\mathbf{x}$*, the optimal facility location with respect to the miniSOS objective is the average location; i.e.,* $Opt = argmin_y sc(y, \mathbf{x}) = \frac{\sum_{i\in N} x_i}{n}$.

The following lemma proves extremely useful in establishing the lower bounds throughout this section. In particular, it helps us relate joint deviations (i.e., coordinated deviations by a subset of agents) to unilateral deviations (i.e., deviations by a single agent). We note that this lemma is a special case of lemma 2.1 from Lu et al. [2010]; it is presented and proved here for completeness.

LEMMA 4.2. *Let* $a, b, c \in \mathbb{R}$ *be three locations such that* $a \le b \le c$*, with at least one strict inequality, and for every* $m \in [n]$*, let* $\mathbf{x}^0$ *(respectively,* $\mathbf{x}^m$*) be a location profile in which* $n-m$ *agents are located at* a*, and* m *agents are located at* c *(resp.,* b*). Let* f *be a randomized mechanism. If* f *is an SP mechanism, then* $E[|c-y^0|] \le E[|c-y^m|]$ *and* $E[|b-y^m|] \le E[|b-y^0|]$*, where* $y^0 \sim f(\mathbf{x}^0)$ *and* $y^m \sim f(\mathbf{x}^m)$.

4.1. Deterministic mechanisms

THEOREM 4.3. *Given a location profile* $\mathbf{x}$*, the mechanism that chooses the median location in* $\mathbf{x}$ *is an SP* 2*-approximation mechanism for the miniSOS objective.*

PROOF. Let μ be a median location in $\mathbf{x}$. It is well known that the median mechanism on a line is truthful (see, e.g., [Procaccia and Tennenholtz 2009]). Therefore, it remains to show that it provides a 2-approximation to the miniSOS objective; formally, we need to show that

$$\sum_{i\in N} |x_i - \mu|^2 \le 2 \sum_{i\in N} |Opt - x_i|^2. \tag{1}$$

Assume without loss of generality that $x_1 \le x_2 \le \ldots \le x_n$. Assume additionally that $\mu \le Opt$ (the proof works analogously for the case in which $\mu \ge Opt$). It is easy to verify that

$$\sum_{i\in N}|x_i-\mu|^2=\sum_{i\in N}|(x_i-Opt)+(Opt-\mu)|^2$$
$$=\sum_{i\in N}\left((x_i-Opt)^2+2(Opt-\mu)(x_i-\frac{Opt+\mu}{2})\right).$$

Therefore, by subtracting $\sum_{i\in N}|Opt-x_i|^2$ from both sides, it remains to prove that

$$\sum_{i\in N}2(Opt-\mu)(x_i-\frac{Opt+\mu}{2})\leq\sum_{i\in N}|Opt-x_i|^2,$$

which is equivalent to showing that

$$\sum_{i\leq\lceil n/2\rceil}2(Opt-\mu)\left((x_i-\frac{Opt+\mu}{2})+(x_{n+1-i}-\frac{Opt+\mu}{2})\right)\leq$$
$$\sum_{i\leq\lceil n/2\rceil}|Opt-x_i|^2+|Opt-x_{n+1-i}|^2.$$

We next show that the last inequality holds piecewise; i.e., for every $i\leq\lceil n/2\rceil$, it holds that

$$2(Opt-\mu)\left((x_i-\frac{Opt+\mu}{2})+(x_{n+1-i}-\frac{Opt+\mu}{2})\right)\leq(Opt-x_i)^2+(Opt-x_{n+1-i})^2.$$

For every $i\leq\lceil n/2\rceil$, it holds that $x_i\leq\mu\leq Opt$; thus $(Opt-\mu)^2\leq(Opt-x_i)^2$, and it suffices to show that

$$2(Opt-\mu)\left(\mu-\frac{Opt+\mu}{2}+x_{n+1-i}-\frac{Opt+\mu}{2})\right)\leq(Opt-\mu)^2+(Opt-x_{n+1-i})^2.$$

It can be easily verified that the above inequality holds if and only if $(2Opt-\mu-x_{n+1-i})^2\geq 0$; the assertion of the theorem follows. □

Notably, this mechanism is a special case of Mechanism PB.

The following theorem shows that factor 2 is tight with respect to deterministic SP mechanisms.

THEOREM 4.4. *Any deterministic truthful mechanism has an approximation ratio of at least* 2 *for the miniSOS objective.*

PROOF. Assume by way of contradiction that there exists a deterministic SP mechanism f which yields a better approximation than 2. Consider a location profile $\mathbf{x}$, in which $\frac{n}{2}$ agents are located at 0, and $\frac{n}{2}$ agents are located at 2, and let $f(\mathbf{x})=p$. Simple calculations show that to achieve a better approximation than 2, it must hold that $p\in(0,2)$ (note that the optimal location is 1, which obtains an SOS cost of n, while the locations 0 or 2 obtains each an SOS cost of $2n$). Now consider a different location profile, denoted $\mathbf{x}^p$, in which $\frac{n}{2}$ agents are located at 0, and $\frac{n}{2}$ agents are located at p. Following the same argument, it must hold that $f(\mathbf{x}^p)\in(0,p)$, and thus $|p-f(\mathbf{x}^p)|>0$. Since $f(\mathbf{x})=p$, we get that $|p-f(\mathbf{x}^p)|>|p-f(\mathbf{x})|$, which implies, by Lemma 4.2, that f is not SP. □

4.2. Randomized mechanisms

A natural candidate of a randomized mechanism to be considered in our context is the *random dictator* (RD) mechanism, which chooses each agent's location with probability $\frac{1}{n}$. This mechanism is SP and is known to provide a $\left(2-\frac{2}{n}\right)$-approximation with respect to the minisum objective function (See [Alon et al. 2009] and [Alon et al. 2010]). The following theorem shows that the RD mechanism provides a 2-approximation for the miniSOS objective. More precisely, for *every* location profile, the RD mechanism yields an SOS cost that is exactly twice the cost of the optimal location. This is established in the following theorem.

THEOREM 4.5. *For every location profile, the RD mechanism yields an SOS cost that is exactly twice the optimal SOS cost.*

As shown in Proposition 3.6, the RD mechanism is a probability distribution over PB mechanisms.

Apparently, the RD mechanism does not perform better than the deterministic median mechanism. Yet, this mechanism turns out to be useful when integrated within a more sophisticated mechanism, as shown below.

MECHANISM 1. *Given $\mathbf{x} \in R^n$, choose the average point with probability $\frac{1}{2}$, and apply the RD mechanism with probability $\frac{1}{2}$ (i.e., for every $i \in N$, x_i is chosen with probability $\frac{1}{2n}$).*

The following theorem establishes the strategyproofness and the approximation ratio provided by Mechanism 1. Note that the PB mechanism formulation allows a straightforward proof of the SP component.

THEOREM 4.6. *Mechanism 1 is an SP* 1.5*-approximation for the SOS objective.*

PROOF. We first prove the approximation factor. Let $g(\mathbf{x})$ denote mechanism 1, and let $avg(\mathbf{x})$ denote the average point. By Observation 4.1, the optimal location with respect to miniSOS is $avg(\mathbf{x})$. We get

$$\frac{sc(g(\mathbf{x}), \mathbf{x})}{Opt} = \frac{\frac{1}{2}RD(\mathbf{x}) + \frac{1}{2}avg(\mathbf{x})}{avg(\mathbf{x})} = 1.5,$$

where the last equation follows by Theorem 4.5.

In order to show the strategyproofness of the mechanism, it suffices to prove that it is an instance of Mechanism PB. Indeed, one can easily verify that this is a special case in which $m = n$, $\mathbf{w}$ is the uniform distribution over $[n]$, and for every $i \in [n]$, $f_i(\mathbf{x}) = x_i$ (i.e., f_i is dictatorship with agent i as the dictator). □

Notably, while approximation in its usual sense looks at the worst-case ratio between the expected cost of the mechanisms solution and the cost of the optimal solution, in this case the 1.5-approximation applies not only in the worst-case notion; rather, this is the exact approximation achieved for every location profile.

Surprisingly, Mechanism 1 provides the best possible approximation; that is, no SP mechanism, randomized or not, can achieve a better approximation ratio than 1.5. This bound is established in the next theorem.

THEOREM 4.7. *Any randomized SP mechanism has an approximation ratio of at least* 1.5 *for the miniSOS objective.*

PROOF. Assume on the contrary that there exists an SP mechanism $f(\mathbf{x})$ and $\epsilon > 0$ such that $f(\mathbf{x})$ yields an approximation ratio of $1.5 - \epsilon$, and let $y \sim f(x)$ (i.e., y is a random variable that is distributed according to $f(x)$). We shall use the following lemma in the proof.

LEMMA 4.8. *There exists some $a \in \mathbb{R}$ and a location profile $\mathbf{x}$ in which $\frac{n}{2}$ agents are located at a and $\frac{n}{2}$ agents are located at $a+4$, such that $E[|y-(a+1)|+|y-(a+3)|] > 3-2\epsilon$.*

PROOF. Consider a location profile $\mathbf{x}^0$, in which $\frac{n}{2}$ agents are located at 0 and $\frac{n}{2}$ agents are located at 4, and let $y^0 \sim f(\mathbf{x}^0)$. If $E[|y^0 - 3| + |y^0 - 1|] > 3 - 2\epsilon$, then we are done. Otherwise, either $E[|y^0 - 1|] \le 1.5 - \epsilon$ or $E[|y^0 - 3|] \le 1.5 - \epsilon$. Assume w.l.o.g. that the latter holds, and consider a location profile $\bar{\mathbf{x}}^0$, in which $\frac{n}{2}$ agents are located at 0 and the rest are located at 3. Let $\bar{y}^0 \sim f(\bar{\mathbf{x}}^0)$. By Lemma 4.2, to preserve truthfulness, it must hold that $E[|\bar{y}^0 - 3|] \le 1.5 - \epsilon$, which implies that $E[|\bar{y}^0 - 0|] \ge 1.5 + \epsilon$.

Consider next a location profile $\mathbf{x}^1$, in which $\frac{n}{2}$ agents are located at -1 and the rest are located at 3, and let $y^1 \sim f(\mathbf{x}^1)$. By Lemma 4.2, to preserve truthfulness, it must hold that $E[|y^1 - 0|] \ge 1.5 + \epsilon$. If $E[|y^1 - 2| + |y^1 - 0|] > 3 - 2\epsilon$, then we are done. Otherwise, it follows that $E[|y^1 - 2|] \le 1.5 - 3\epsilon$.

We continue iterating such that in iteration $j = 1, 2, \ldots$, a profile $\mathbf{x}^j$ is considered, in which half the agents are located at $-j$ and half are located at $4 - j$, and for every profile $\mathbf{x}^j$, we denote by y^j the random variable distributed according to $f(\mathbf{x}^j)$. We show that there exists some j for which $E[|y^j - (1-j)| + |y^j - (3-j)|] > 3 - 2\epsilon$; the assertion of the lemma then follows by substituting $a = -j$. It remains to prove the last inequality. Indeed, by repeatedly applying Lemma 4.2 for every j, we get that

$$\text{if } E[|y^j - (1-j)| + |y^j - (3-j)|] \le 3 - 2\epsilon, \text{ then } E[|y^j - (3-j)|] \le 1.5 - \epsilon(2j+1). \quad (2)$$

But since for $j > \frac{1.5-\epsilon}{2\epsilon}$, it holds that $1.5 - \epsilon(2j+1) < 0$, it must hold that $E[|y^j - (3-j)|] > 1.5 - \epsilon(2j+1)$, which, by Equation 2 implies that $E[|y^j - (1-j)| + |y^j - (3-j)|] > 3 - 2\epsilon$. It follows that the profile $\mathbf{x}^j$ satisfies the conditions of the lemma, and the proof follows. □

With this lemma, we are ready to prove the theorem. Let $\mathbf{x}$ be a location profile that satisfies the conditions of Lemma 4.8, and assume w.l.o.g. that $\mathbf{x}$ is a profile in which half the agents are located at 0 and half at 4. By the last lemma, it holds that

$$E[|y - 1| + |y - 3|] > 3 - 2\epsilon, \quad (3)$$

where $y \sim f(\mathbf{x})$. For ease of presentation, let $p = Pr(|y - 2| \le 1)$ and let $z = E[|y - 2| : |y - 2| > 1]$. It holds that

$$\begin{aligned} E[|y-1| + |y-3|] &= E\left[|y-1| + |y-3| : |y-2| > 1\right](1-p) \\ &\quad + E\left[|y-1| + |y-3| : |y-2| \le 1\right]p \\ &= 2z(1-p) + 2p. \end{aligned}$$

Therefore, by Equation (3), it follows that $2z(1-p) + 2p > 3 - 2\epsilon$. Since $z > 1$ by definition, it follows that

$$p < 1 - \frac{1 - 2\epsilon}{2z - 2}. \quad (4)$$

We now turn to calculate the SOS cost of the profile $\mathbf{x}$ induced by the mechanism $f(\mathbf{x})$. It holds that

$$\begin{aligned} E[sc(y, \mathbf{x})] &= E[sc(y, \mathbf{x}) : |y - 2| \le 1]p + E[sc(y, \mathbf{x}) : |y - 2| > 1](1-p) \\ &\ge sc(E[y : |y - 2| \le 1], \mathbf{x})p + E[sc(y, \mathbf{x}) : |y - 2| > 1](1-p), \end{aligned} \quad (5)$$

where the last inequality follows from Jensen's inequality. Since 2 is the optimal location in the profile $\mathbf{x}$, it holds that

$$sc(E[y : |y - 2| \le 1], \mathbf{x}) \ge sc(2, \mathbf{x}) = 4n. \quad (6)$$

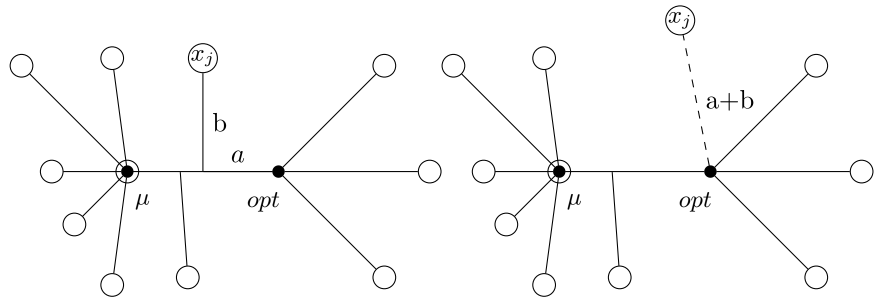

Fig. 1. An illustration of the iterative process described in the proof of Theorem 5.1. The transformed graph is illustrated on the right, where the dashed line represents the new edge.

We also have that

$$\begin{aligned} E[sc(y,\mathbf{x}) : |y-2| > 1] &= E[\frac{n}{2}(y^2 + (y-4)^2) : |y-2| > 1] \\ &= \frac{n}{2}(8 + 2E[(y-2)^2 : |y-2| > 1]) \\ &\geq \frac{n}{2}(8 + 2E^2[|y-2| : |y-2| > 1]) \\ &= \frac{n}{2}(8 + 2z^2), \end{aligned} \tag{7}$$

where the last inequality follows from Jensen's inequality. By Substituting (6) and (7) in (5), we get that $E[sc(y,\mathbf{x})] \geq \frac{n}{2}(8 + 2z^2 - 2z^2p)$. It, therefore, follows by (4) that

$$E[sc(y,\mathbf{x})] > \frac{n}{2}(8 + 2z^2 - 2z^2(1 - \frac{1-2\epsilon}{2z-2})) = 4n + \frac{nz^2(1-2\epsilon)}{2z-2}. \tag{8}$$

It is easy to verify that this function attains its minimum at $z = 2$, with a value of $6n - 4n\epsilon$. Thus, $E[sc(y,\mathbf{x})] > 6n - 4n\epsilon$. The optimal solution is to locate the facility at 2, which yields an SOS cost of $4n$. We get that $\frac{E[sc(y,\mathbf{x})]}{Opt} > \frac{6n-4n\epsilon}{4n} = 1.5 - \epsilon$, and a contradiction is reached. The assertion of the theorem follows. □

5. SP MECHANISMS ON A TREE

In this section, we study the miniSOS objective with respect to locating the facility on a tree. In the deterministic case, we show that the median of a tree provides a 2-approximation for the miniSOS objective, and show that no SP deterministic mechanism can achieve a better ratio with respect to miniSOS. In the randomized case, we construct an instance of Mechanism PB, which obtains a 1.83-approximation.

5.1. Deterministic mechanisms

It is well known that the mechanism that chooses a median of a tree is SP (see, e.g., Alon et al. [2009]). The following theorem establishes that a median also gives a 2-approximation for the miniSOS objective, which is tight, according to Theorem 4.4.

THEOREM 5.1. *The median of a tree is an SP 2-approximation mechanism for the miniSOS objective.*

A sketch of the proof follows.

PROOF SKETCH. Given an instance $(G,\mathbf{x})$, we iteratively transform it, in a way that can only make the approximation ratio obtained by the median worse, and eventually prove the desired approximation ratio on the final instance. Let μ and Opt denote the respective median and optimal location in the original profile. The iterative process proceeds as follows. As long as there exists an agent j such that x_j's subtree is rooted at the open interval $path(\mu, Opt)$, pick such an agent j, create a new edge of length $d(x_j, Opt)$, rooted at Opt, and locate x_j at its tip (see illustration in Figure 1). It can be proved that the median and the optimal location did not change as a result of this transformation, and that the optimal cost did not change either. The cost of the median, however, can only increase. Therefore, the approximation ratio can only get worse by each transformation. Upon termination of this process, we prove the desired approximation ratio on the final instance by reducing it to the deterministic scenario on a line and applying Theorem 4.3. □

5.2. Randomized mechanisms

In this section we present an instance of Mechanism PB, which obtains 1.83-approximation for trees, with respect to the miniSOS objective. Our randomized mechanism uses the following deterministic mechanism as a building block.

Mechanism **dictatorial-generalized-median** *(DGM).* Mechanism DGM receives as parameters an index $i \in [n]$, and a fraction $q \in (1/2, 1]$. The facility location is chosen deterministically, with respect to i and q, as follows. Fix the point x_i as the root of the tree, and denote the current location a. Then, as long as there exists a subtree in $T(G, a)$ that contains at least fraction q of the agents (this is well defined, since $q > 1/2$), smoothly move down this subtree. Finally, when we reach a point where it is not possible to move closer to at least fraction q of the agents by continuing downwards, stop and return the current location.

As an illustration, consider Mechanism DGM applied on the line with $q = \frac{3}{4}$. Let ℓ be the agent that is positioned $\lceil \frac{n}{4} \rceil$ agents from the left, and r be the agent that is positioned $\lceil \frac{n}{4} \rceil$ agents from the right. Mechanism DGM with $q = \frac{3}{4}$ operates as follows: If the agent i (that is specified as part of the mechanism) is located to the left of ℓ, the mechanism proceeds to the right up to the location of agent ℓ (where it can no longer proceeds toward more than $q = \frac{3}{4}$ of the agents), and locate the facility at ℓ. Similarly, if the agent i is located to the right of r, the mechanism proceeds to the left up to the location of agent r. For any other agent i (i.e., i such that $x_i \in [\ell, r]$), the mechanism cannot proceed to any direction and therefore locates the facility at x_i.

This mechanism is a boomerang mechanism, as asserted by the following proposition[7].

Proposition 5.2. *Mechanism DGM is a boomerang mechanism.*

With this we are ready to introduce our randomized mechanism.

Mechanism **randomized DGM**. Mechanism randomized DGM receives as a parameter a fraction $q \in (1/2, 2/3]$, and applies Mechanism PB with the following parameters: $m = n$, and for every $i \in [n]$, $w_i = 1/n$ and $f_i(\mathbf{x})$ is Mechanism DGM with parameters i and q.

The following lemma establishes an important property of mechanism randomized DGM that will be used in its analysis.

Lemma 5.3. *For every tree G, the points $f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})$, calculated by Mechanism randomized DGM, are located on a single path.*

With Lemma 5.3 at hand, it is easy to describe mechanism randomized DGM more intuitively: Given a fraction $q \in (1/2, 2/3]$, for each agent i, let y_i denote the location chosen by Mechanism DGM with index i. Mechanism randomized DGM then chooses a random dictator among $y_1, y_2, \ldots, y_n$ with probability $\frac{1}{2}$, and their average location with probability $\frac{1}{2}$ (where the average is well defined by Lemma 5.3.

[7]Note that Mechanism DGM is not a boomerang mechanism for $q \leq 1/2$, as a beneficial misreport can exist, depending on the tie-breaking rule.

In the remainder of this section, we shall use the notation suggested above, i.e., for every $i \in [n]$, let $y_i = f_i(\mathbf{x})$. In addition, we assume that w.l.o.g, y_1 and y_n are located on the edges of the path created by $\mathbf{y}$, and we let $avg(\mathbf{y})$ denote the average location on $path(y_1, y_n)$.

The main result of this section establishes that Mechanism randomized DGM, with $q = 2/3$, obtains an approximation ratio of 1.83 for any tree network.

THEOREM 5.4. *Let G and $\mathbf{x}$ be a tree and a location profile, respectively. Mechanism randomized DGM with $q = 2/3$ obtains an approximation ratio of 1.83 with respect to the miniSOS objective.*

Before presenting the proof sketch of the theorem, we observe that it can be assumed w.l.o.g that $y_1 \neq y_n$, due to the following lemma.

LEMMA 5.5. *If $y_1 = y_2 = \ldots = y_n$, then the approximation ratio obtained by Mechanism randomized DGM is at most 1.5.*

We are now ready to present an overview of the proof of Theorem 5.4. The proof of this theorem proceeds through several lemmata, which correspond to various transformations that are performed in order to establish the desired approximation ratio. Due to space limitation, only a sketch of the proof is provided, along with some graphical illustrations. The full proof, including all the required lemmata, is deferred to the full version.

PROOF SKETCH. Following Lemma 5.5, we assume that $y_1 \neq y_n$. Additionally, we scale the graph so that $d(y_1, y_n) = 1$, and let $\bar{Opt}$ denote the closest location on $path(y_1, y_n)$ to Opt, i.e., $\bar{Opt} = argmin_{l \in path(y_1,y_n)} d(Opt, l)$.

The proof is established as follows. Given a graph G and a location profile $\mathbf{x}$, we perform various transformations, in such a way that the approximation ratio obtained by the mechanism could only get worse with every transformation. Eventually, we prove that the mechanism provides a 1.83-approximation ratio on the final graph and location profile, which implies the same upper bound on the original instance.

More specifically, given an instance $(G, \mathbf{x})$, we proceed as follows:

We denote by p the location of $\bar{opt}$ in the original $(G, \mathbf{x})$, and we fix it throughout the process, i.e., p shall stay fixed even when $\bar{opt}$ moves due to a transformation. In the first stage, we transform the graph repeatedly in the following way. For each agent i such that y_i is on the open $path(y_1, y_n)$ but x_i is not, we introduce two alternative transformations to the graph in which only i is relocated, while preserving her distance from $\bar{opt}$. We then prove that in at least one of them the approximation ratio gets worse. Assuming that y_i is between y_1 and $\bar{opt}$ (the proof works analogously for the case in which y_i is between $\bar{opt}$ and y_n), the alternatives are either to locate i at the tip of a new edge, rooted at $\bar{opt}$; or to locate i on a path that coincides with $path(y_1, \bar{opt})$ (if it exceeds $d(y_1, \bar{opt})$, we create a new edge rooted at y_1, to continue $path(y_1, \bar{opt})$ as necessary). We apply this lemma repeatedly until all of the agents that are not located on $path(y_1, y_n)$ are rooted at either y_1, y_n, or $\bar{Opt}$, with respect to $path(y_1, y_n)$ (see Figure 2).

In the second stage, we deal with the agents that are rooted at y_1 and y_n, with respect to $path(y_1, y_n)$. We assume that $avg(\mathbf{y})$ is on $path(y_1, \bar{Opt})$ (the proof follows analogously if $avg(\mathbf{y})$ is on $path(\bar{Opt}, y_n)$). We show that relocating the agents that are rooted at y_1 to y_1 (see Figure 3) would only worsen the approximation ratio. We then show that relocating all of the agents rooted at y_n, except for those that are in the same subtree as Opt (i.e., all agents i such that $y_i = y_n$, except for agents i such that $path(Opt, y_n) \cap path(x_i, y_n) \neq y_n$), on their own edges, equally distanced from y_n, at a distance of their average distance from y_n (see Figure 4), would only worsen the approximation ratio.

At this point, if Opt is rooted at y_n (with respect to $path(y_1, y_n)$), and p is not at y_n, we repeat the first two stages. We note that the assumption of $avg(\mathbf{y})$ being on $path(y_1, \bar{Opt})$ would still hold, as $\bar{Opt}$ would stay at y_n. We also note that this action would result in a

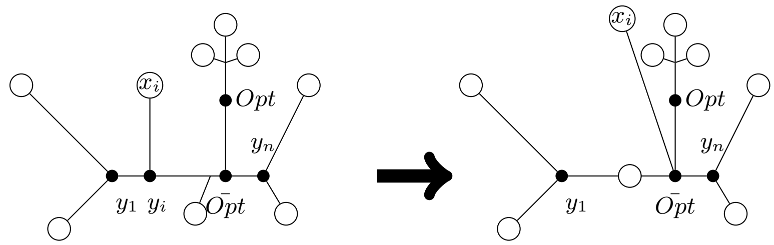

Fig. 2. The first stage assures that all the agents that are not located on $path(y_1, y_n)$ are rooted at either y_1, y_n, or $\bar{Opt}$.

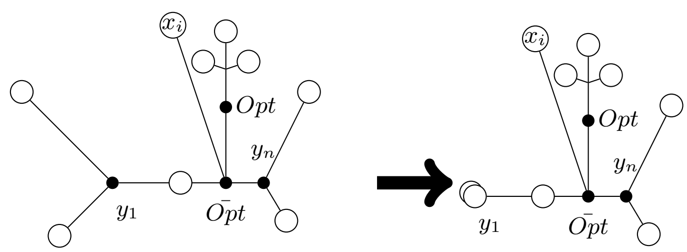

Fig. 3. The second stage - assuring that all agents that are rooted at y_1 are in fact located at y_1.

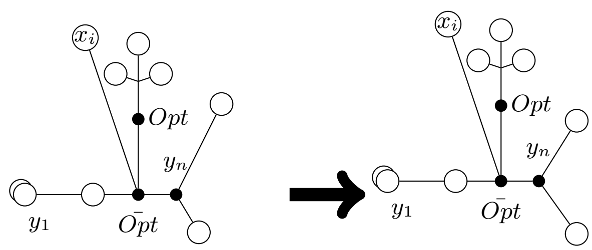

Fig. 4. The second stage - averaging the distance of the agents that are rooted at y_n.

graph in which $\bar{opt}$ is located at y_n, at least $|N|/3$ agents are located at y_1, and all of the other agents are either on $path(y_1, y_n)$, or rooted at y_n.

We now distinguish between three cases:

— If Opt is located at y_n, it can be shown that all of the agents are either at y_1, scattered on $path(y_1, y_n)$, or rooted at y_n, equidistant from y_n and on different edges. In this case we establish a $1\frac{1}{2}$-approximation.
— If $\bar{Opt}$ is located at y_n, but Opt is not located at y_n (i.e., opt is in a subtree rooted at y_n), we transform the graph in the following way. Consider the agents that are rooted at y_n and are in the same subtree of y_n as Opt. Let σ be their average distance from y_n. Create a new edge of length σ, rooted at y_n, and place all of them at its tip. We show that the transformation could only worsen the ratio, and we then show an approximation ratio of 1.83 on the obtained graph.
— Otherwise (meaning $\bar{Opt}$ is on the open interval $path(y_1, y_n)$, and so is p), we transform the graph as follows. Consider the agents that are rooted at p, and let σ be their average distance from p. Locate them on the tip of a new edge of length σ, rooted at p (see Figure 5). We show again that this transformation can only worsen the ratio, and an approximation ratio of 1.83 is obtained in the resulting graph.

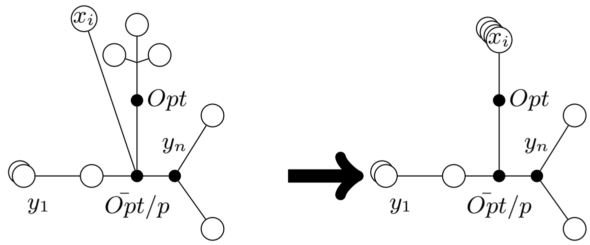

Fig. 5. Averaging the distance of the agents that are rooted at p, while locating them on the same edge.

Finally, by observing that $\bar{Opt}$ cannot be located at y_1, we conclude that these three cases exhaust all the possibilities for Opt's location. In conclusion, since all the transformations are shown to only worsen the approximation ratio, the 1.83-approximation that is obtained with respect to the final instance, imposes the same upper bound on the original one, and the assertion of the theorem follows.

□

6. DISCUSSION AND OPEN PROBLEMS

We study the problem of strategyproof facility location on tree networks. Our two main contributions are (i) the introduction of a sufficient condition for strategyproofness, which essentially provides a broad family of randomized SP mechanisms, and (ii) an analysis of the approximation ratio that can be obtained by an SP mechanism with respect to the miniSOS objective function.

Our study leaves many questions for future research. We believe that the foundations laid by our study can be used in exploring these directions. Some of the most intriguing open problems are the following:

— Providing a full characterization for SP randomized mechanisms on a tree. The sufficient condition presented in this paper might be a good starting point for the characterization, although it is possible that a different avenue should be taken with respect to full characterization.
— Closing the approximation gap (between 1.83 and 1.5) for randomized SP mechanisms on a tree with respect to the miniSOS objective.
— Analyzing the approximation ratios with respect to the miniSOS function for additional network topologies (such as a cycle) and, more ambitiously, for general networks.
— Extending this study to additional individual cost functions. For example, in applications in which agents are more sensitive to distances within the range of high distances, a convex cost function seems plausible (a possible example might be the sensitivity of users to the speed of an Internet connection).
— Finally, the three different social functions that have been studied thus far can be considered as special cases of the ℓ-norm distance, with minisum, miniSOS, and minimax corresponding to the 1-norm, 2-norm, and ∞-norm, respectively. It is apparent that while for the minisum function, the optimal location can be obtained in an SP mechanism, this is not feasible for either 2- or ∞-norms, and the same approximation bounds apply in both cases. Generalizing this result to any ℓ-norm is an additional stimulating direction.

REFERENCES

Alon, N., Feldman, M., Procaccia, A. D., and Tennenholtz, M. 2009. Strategyproof approximation mechanisms for location on networks. *CoRR abs/0907.2049*.

Alon, N., Feldman, M., Procaccia, A. D., and Tennenholtz, M. 2010. Strategyproof approximation of the minimax on networks. *Math. Oper. Res. 35,* 3, 513–526.

Ashlagi, I., Fischer, F., Kash, I., and Procaccia, A. D. 2010. Mix and match. In *Proceedings of the 11th ACM conference on Electronic commerce.* EC '10. ACM, New York, NY, USA, 305–314.

Bandelt, H.-J. 1985. Networks with condorcet solutions. *European Journal of Operational Research 20,* 3, 314–326.

Bandelt, H. J. and Labbé, M. 1986. How bad can a voting location be. *Social Choice and Welfare 3*, 125–145. 10.1007/BF00435663.

Black, D. 1958. *The theory of committees and elections.* University Press, Cambridge.

Border, K. C. and Jordan, J. S. 1983. Straightforward elections, unanimity and phantom voters. *Review of Economic Studies 50,* 1, 153–170.

Dekel, O., Fischer, F., and Procaccia, A. D. 2010. Incentive compatible regression learning. *Journal of Computer and System Sciences 76,* 8, 759 – 777.

Dughmi, S. and Ghosh, A. 2010. Truthful assignment without money. *CoRR abs/1001.0436.*

Ehlers, L., Peters, H., and Storcken, T. 2002. Strategy-proof probabilistic decision schemes for one-dimensional single-peaked preferences. *Journal of Economic Theory 105,* 2, 408–434.

Fotakis, D. and Tzamos, C. 2010. Winner-imposing strategyproof mechanisms for multiple facility location games. In *Proceedings of the 6th international conference on Internet and network economics.* WINE'10. Springer-Verlag, Berlin, Heidelberg, 234–245.

Handler, G. and Mirchandani, P. 1979. *Location on networks: theory and algorithms.* MIT Press series in signal processing, optimization, and control. MIT Press.

Holzman, R. 1990. An axiomatic approach to location on networks. *Math. Oper. Res. 15*, 553–563.

Labbe, M. 1985. Outcomes of voting and planning in single facility location problems. *European Journal of Operational Research 20,* 3, 299–313.

Lu, P., Sun, X., Wang, Y., and Zhu, Z. A. 2010. Asymptotically optimal strategyproof mechanisms for two-facility games. In *Proceedings of the 11th ACM conference on Electronic commerce.* EC '10. ACM, New York, NY, USA, 315–324.

Lu, P., Wang, Y., and Zhou, Y. 2009. Tighter bounds for facility games. In *Internet and Network Economics, 5th International Workshop, WINE 2009, Rome, Italy, December 14-18, 2009. Proceedings*, S. Leonardi, Ed. Lecture Notes in Computer Science Series, vol. 5929. Springer, 137–148.

Marsh, M. T. and Schilling, D. A. 1994. Equity measurement in facility location analysis: A review and framework. *European Journal of Operational Research 74,* 1, 1–17.

Moulin, H. 1980. On strategy-proofness and single peakedness. *Public Choice 35,* 4, 437–455.

Procaccia, A. D. and Tennenholtz, M. 2009. Approximate mechanism design without money. In *Proceedings of the 10th ACM conference on Electronic commerce.* EC '09. ACM, New York, NY, USA, 177–186.

Schummer, J. and Vohra, R. V. 2002. Strategy-proof location on a network. *Journal of Economic Theory 104,* 2, 405–428.