

Strategyproof Approximation Mechanisms for Location on Networks

Noga Alon* Michal Feldman† Ariel D. Procaccia‡ Moshe Tennenholtz §

Abstract

We consider the problem of locating a facility on a network, represented by a graph. A set of strategic agents have different ideal locations for the facility; the cost of an agent is the distance between its ideal location and the facility. A *mechanism* maps the locations reported by the agents to the location of the facility. Specifically, we are interested in social choice mechanisms that do not utilize payments. We wish to design mechanisms that are *strategyproof*, in the sense that agents can never benefit by lying, or, even better, *group strategyproof*, in the sense that a coalition of agents cannot all benefit by lying. At the same time, our mechanisms must provide a small approximation ratio with respect to one of two optimization targets: the social cost or the maximum cost.

We give an almost complete characterization of the feasible truthful approximation ratio under both target functions, deterministic and randomized mechanisms, and with respect to different network topologies. Our main results are: We show that a simple randomized mechanism is group strategyproof and gives a $(2 - 2/n)$ -approximation for the social cost, where n is the number of agents, when the network is a circle (known as a *ring* in the case of computer networks); we design a novel “hybrid” strategyproof randomized mechanism that provides a tight approximation ratio of $3/2$ for the maximum cost when the network is a circle; and we show that no randomized SP mechanism can provide an approximation ratio better than $2 - o(1)$ to the maximum cost even when the network is a tree, thereby matching a trivial upper bound of two.

*Microsoft Israel R&D Center, 13 Shenkar Street, Herzeliya 46725, Israel, and Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv, 69978, Israel, Email: nogaa@tau.ac.il. Research supported in part by a USA Israeli BSF grant, by a grant from the Israel Science Foundation, by an ERC advanced grant and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

†Microsoft Israel R&D Center, 13 Shenkar Street, Herzeliya 46725, Israel, and School of Business Administration and Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. Email: mfeldman@huji.ac.il

‡Microsoft Israel R&D Center, 13 Shenkar Street, Herzeliya 46725, Israel. Email: arielpro@cs.huji.ac.il.

§Microsoft Israel R&D Center, 13 Shenkar Street, Herzeliya 46725, Israel, and Technion, IIT, Haifa 32000, Israel. Email: moshet@microsoft.com

1 Introduction

We consider a setting which consists of a network (represented by a graph) and a set of strategic agents. We would like to locate a facility on the network, in a way that is based on the preferences of the agents. However, the agents may disagree on the ideal location for the facility; in particular, the *cost* of an agent given a facility location is the distance of the facility from the agent's ideal location.

This abstract setting has many natural interpretations. For example, a typical social choice scenario concerns a network of roads in a town. The mayor would like to choose a location for a public facility, such as a library or a fire station. Some citizens would like the facility to be close to their homes whereas others may wish it to be adjacent to their workplace, but they all have (presumably different) ideal locations.

A second, prominent example is obtained by supposing that the network is a telecommunications network, such as a local computer network or the Internet. In this case, the agents are the network users or service providers, and the facility can be, e.g., a filesharing server or a router. This interpretation also motivates attention to specific, common network topologies such as tree networks (also known as hierarchical networks), star networks (which are, graph-theoretically speaking, a special case of trees), and ring networks.

Finally, Schummer and Vohra [13] point out that the network can also be *virtual* rather than physical. For instance, a group of people may wish to schedule a daily meeting or task (such as a file backup). In this case, the network is simply a single cycle around the face of a (24 hour) clock; a person whose ideal time is 23:59 should be indifferent between 23:57 and 00:01.

A *mechanism* in the above setting is a function that receives the ideal locations of the agents as input, and returns the facility location. Our game-theoretic goal is to design mechanisms that are truthful or *strategyproof* (SP), in the sense that an agent cannot benefit by misreporting its ideal location, regardless of the reports of the other agents. Moreover, we would like our mechanisms to be approximately optimal with respect to a target function, where approximation is defined in the usual sense by looking at the worst-case ratio between the cost of the mechanism's solution and the cost of the optimal solution. Specifically, we are interested in two target functions: the social cost (the sum of agents' distances to the facility) and the maximum cost (the largest distance to the facility). In the road network example minimizing the social cost makes more sense if the facility is a library, whereas one would wish to minimize the maximum cost if the facility were a fire station.

Crucially, if one allows the mechanism to make payments, that is, if money is available, then optimal truthful mechanisms can be obtained by using generic mechanism design techniques, such as augmenting the optimal solution with Vickrey-Clarke-Groves (VCG) payments (see, e.g., [9]). However, in many settings money cannot be used to induce truthfulness, due to ethical or legal considerations (see, e.g., [14]). Moreover, in Internet (or, in general, telecommunications) environments making payments is often technically infeasible due to security and banking issues. Finally, mechanisms like VCG are known to suffer from a host of disadvantages and paradoxes [11]. Therefore, we will be interested in moneyless mechanisms whose very definition precludes payments. Such mechanisms commonly appear in the social choice literature, where they are sometimes known as *social choice rules*. It is important to note that we do not use approximation to circumvent computational hardness (as our optimization problems are tractable), but rather approximation is employed to obtain strategyproofness (at the expense of the optimality of the solution) without resorting to payments.

Related work. The agenda of *approximate mechanism design without money*, namely the study of *moneyless* SP approximation mechanisms for optimization problems, was recently explicitly advocated by Pro-

caccia and Tennenholtz [10], but can be traced back to work on incentive compatible learning by Dekel et al. [4]. The basic setting of Procaccia and Tennenholtz is a very special case of ours, where the network is, very simply, a line. It is trivial that if the target function is the social cost, there is an optimal (moneyless) SP mechanism. However, if one wishes to minimize the maximum cost, the best deterministic SP mechanism has an approximation ratio of two, whereas the best randomized SP mechanism has a ratio of $3/2$. These straightforward results are summarized in the left column of Table 1. Procaccia and Tennenholtz focus on two extensions of the basic setting: location of two facilities on a line, and location of one facility on a line when each agent controls multiple points. All their results hold only when the agents are located on a line, and they do not consider (more general) networks.

The model investigated by Schummer and Vohra [13] is identical to ours; their goal is to identify the deterministic SP mechanisms in this setting. Schummer and Vohra give a characterization of all the SP mechanisms when the underlying network is a tree. Furthermore, the authors demonstrate that if the network contains a cycle, any SP mechanism is almost dictatorial, in the sense that one fixed agent dictates the location of the facility. The exact notion of dictatorship is discussed in the sequel. This result is analogous to the celebrated Gibbard-Satterthwaite impossibility theorem [5, 12], but holds for the case where the preferences of the agents are restricted by the topology of the network. Note that Schummer and Vohra do not consider optimization problems and approximation, nor do they study randomized mechanisms.

A significant body of work (see, e.g., [6, 7, 3]) deals with a model practically identical to ours, and considers several variations on the following question: how bad can a Condorcet point be in terms of its social cost? A Condorcet point is a location in a network that is preferred by more than half the agents to any other location. The quality of the Condorcet point is measured by bounding the ratio between its social cost and the social cost of the optimal location. In this respect the results are reminiscent of approximation, but the approach is descriptive rather than algorithmic. In addition, this line of work does not directly deal with incentives.

Taking the algorithmic perspective, our work is related to the literature on approximation algorithms for k -median and k -center problems (see, e.g., [2, 1]). However, these problems are tractable when only one facility must be located (i.e., $k = 1$), and this is the case we deal with. As mentioned above, the complexity of our problem (and, hence, the need for approximation) stems from game-theoretic considerations rather than the combinatorial richness of the problem.

Our results and techniques. This paper is the first to study SP (moneyless) approximation mechanisms for facility location in networks. More generally, it is also a very early, and in our opinion the most technically significant, instance of approximate mechanism design without money. Our almost complete characterization of the feasible SP approximation bounds has given rise to three major results; their proofs are elementary but rather complicated and involve some subtle combinatorial and probabilistic ideas.

Our results are summarized in Table 1. We seek SP mechanisms, but in some cases succeed in obtaining even stronger guarantees by showing *group strategyproofness (GSP)*; a mechanism is GSP if whenever a coalition of agents lies, at least one of the members of the coalition does not gain from the deviation.

The upper part of the table contains our results regarding minimization of the social cost, given in Section 3. For the case where the network is a line, there is a straightforward optimal deterministic GSP mechanism. In Section 3.1, we observe that a generalization of this mechanism is GSP and optimal when the network is a tree.

On the other hand, when the network is a circle (e.g., a ring network or the face of a clock), it follows directly from the results of Schummer and Vohra [13] that no deterministic SP Mechanism can obtain an approximation ratio better than $\Omega(n)$. However, randomization turns out to be quite powerful in this context.

		Line [10]	Tree	Circle	General
SC	Det	UB 1 GSP	UB 1 GSP	LB $\Omega(n)$ SP [13]	LB $\Omega(n)$ SP [13]
	Ran	UB 1 GSP	UB 1 GSP	UB $2 - \frac{2}{n}$ GSP (Thm 3.1+3.3) LB open	UB $2 - \frac{2}{n}$ SP (Thm 3.1) LB open
MC	Det	UB 2 GSP LB 2 SP	UB 2 GSP LB 2 SP	UB 2 GSP LB 2 SP	UB 2 GSP LB 2 SP
	Ran	UB 3/2 GSP LB 3/2 SP	UB $2 - \frac{2}{n+2}$ SP LB $2 - o(1)$ SP (Thm 4.2)	UB 3/2 SP (Thm 4.1) LB 3/2 SP	UB 2 GSP LB $2 - o(1)$ SP (Thm 4.2)

Table 1: Summary of our results. The columns correspond to the graph topology, whereas the rows correspond to either social cost (SC) or maximum cost (MC). The rows are further divided into deterministic mechanisms (Det) or randomized mechanisms (Ran). The cells contain the bounds, where UB and LB stand for upper bound and lower bound, respectively, while SP and GSP stand for strategyproof and group strategyproof, respectively (ideally the upper bound is GSP while the lower bound is SP). The number of agents is n . Cells (or columns) with a citation contain results that either appear in, or follow directly from, previous papers. Cells with a theorem number in parentheses contain our main results.

Mechanism 1 simply chooses an agent at random and returns its ideal location. We observe that this trivial mechanism has an approximation ratio of $2 - 2/n$, where n is the number of agents, for the social welfare with respect to any network topology. Our main result in Section 3 is that Mechanism 1 is GSP when the network is a circle. We show that this result can be explicitly stated as follows (with S^1 denoting the unit circle); it may be of interest independently of its game-theoretic interpretation.

Theorem 3.3. Let $x_1, \dots, x_n, y_1, \dots, y_n \in S^1$, and denote the distance on S^1 between $x \in S^1$ and $y \in S^1$ by $d(x, y)$. Then there exists $i \in \{1, \dots, n\}$ such that

$$\sum_{k=1}^n d(x_i, x_k) \leq \sum_{k=1}^n d(x_i, y_k) .$$

The combinatorial heart of the theorem’s proof is the identification of *nearly-antipodal pairs*: pairs of points $\langle x_i, x_j \rangle$ such that there are no other points x_k in the arc between x_i and the antipodal point of x_j , and vice versa. Our proof establishes that there exists such a pair that do not both gain from the deviation, that is, the above equation holds with respect to one of its two members.

In Section 4, we investigate SP mechanisms for minimization of the maximum cost. We observe that even a dictatorship gives us a deterministic GSP 2-approximation mechanism. Moreover, this bound is tight with respect to deterministic mechanisms, even if the network is a line.

In Section 4.1, we deal with randomized mechanisms for the case where the network is a circle, and present Mechanism 2. This mechanism works by combining two mechanisms: one is applied when all the agents are located on one semicircle, whereas the other is applied when the agents are not located on one semicircle. We have the following theorem.

Theorem 4.1. Assume that the network is a circle. Then Mechanism 2 is an SP 3/2-approximation mechanism for the maximum cost.

The result matches an SP lower bound of 3/2. Although the theorem’s proof is quite lengthy, we suggest that the truly striking aspect of this result is the mechanism itself and its strategyproofness. Indeed, the mechanism seems to be a coarse hybridization of two mechanisms, where the combination is required for

achieving the desired approximation ratio. Although each of the two mechanisms is SP in its own right, it seems to us quite extraordinary that the combined mechanism is SP as well.

As it turns out, when we wish to minimize the maximum cost and the network is a tree, randomization cannot significantly help us. Indeed, even though we show that there is a randomized SP mechanism with an approximation ratio of $2 - 2/(n + 2)$, we establish the following lower bound, which is our third and final major result.

Theorem 4.2. Let there be n agents. Then there exists a tree network such that no SP randomized mechanism can have an approximation ratio that is smaller than $2 - \mathcal{O}\left(\frac{1}{2^{\sqrt{\log n}}}\right)$ for the maximum cost.

It is interesting to note that, while with respect to minimization of the social cost there is an optimal SP mechanism on trees but not on circles, when the target function is the maximum cost trees are as hard as general graphs (the optimal ratio is two), whereas on a circle a better SP approximation ratio is feasible.

Open problems. There is still a gap with respect to the optimal ratio achievable by randomized SP mechanisms on a circle, when the target is the social cost. We establish a GSP upper bound of $2 - 2/n$, and it is possible to show a lower bound of $1 + \epsilon$ for a small constant $\epsilon > 0$. We conjecture that there is an SP lower bound of $2 - o(1)$ on a circle. Moreover, while Mechanism 1 is GSP on a circle, it is not GSP in general; we conjecture that there is a lower bound of $\Omega(n)$ with respect to GSP mechanisms in general networks.

A second, small gap has to do with randomized GSP mechanisms on a circle when the target is the maximum cost. Mechanism 2 gives an SP upper bound of $3/2$. However, the mechanism is not GSP, a fact that is not immediately apparent. The question of whether it is possible to achieve a randomized GSP upper bound of $3/2$ on a circle remains open.

2 Preliminaries

We use the model of Schummer and Vohra [13]. Let $N = \{1, \dots, n\}$ be the set of agents. The network is represented by a graph G , formalized as follows. The graph is a closed, connected subset of Euclidean space $G \subset \mathbb{R}^k$. The graph is composed of a finite number of closed curves of finite length, known as the *edges*. The extremities of the curves are known as *vertices*. The agents and the facility may be located anywhere on G .

The reader might feel that the traditional, discrete model of graphs is more appropriate. In addition, other related papers (e.g., [7]) consider a similar continuous model, but allow the agents to be located only on the vertices (whereas the facility can be located anywhere). Crucially, all our results, or slight variations thereof, hold under both these alternative models as well.

The *distance* between two points $x, y \in G$, denoted $d(x, y)$, is the length of the minimum-length path between x and y , where a *path* is a minimal connected subset of G that contains x and y . The center of the path between x and y is denoted $\text{cen}(x, y)$, that is, it is a point z on the path such that $d(x, z) = d(y, z)$. We will also use this notation to denote the center of an interval in \mathbb{R} .

A *cycle* in G is defined to be the union of two paths such that their intersection is equal to the set of both their endpoints. A graph that does not contain cycles is called a *tree*.

We shall be especially interested in the graph that is a single cycle; we refer to such a graph as a *circle*. Given that G is a circle, we denote the shorter open arc between $x, y \in G$ by (x, y) , and the shorter closed arc between x and y by $[x, y]$.¹ For every $x \in G$, we denote by \hat{x} the antipodal point of x on G , that is, the

¹If x and y are antipodal these arcs are ambiguously defined, and when this is problematic we specify to which arc we are referring.

diametrically opposite point. For two points $x, y \in G$, and in the context of an arc of length less than half the circumference of G , we denote the “clockwise operator” by \succeq , and its strong version by \succ ; specifically, $x \succeq y$ means that x is clockwise of y on the circle. We believe that this operator is completely intuitive and requires no formal definition, but in case of need the reader may find such a definition in [13].

Each agent $i \in N$ has an (ideal) *location* $x_i \in G$. The collection $\mathbf{x} = \langle x_1, \dots, x_n \rangle \in G^n$ is referred to as the *location profile*.

A *deterministic mechanism* is a function $f : G^n \rightarrow G$ that maps the reported locations of the agents to the location of a *facility*. When the facility is located at $y \in G$, the cost of agent i is simply the distance between x_i and y :

$$\text{cost}(y, x_i) = d(x_i, y) .$$

A *randomized mechanism* is a function $f : G^n \rightarrow \Delta(G)$, i.e., it maps location profiles to probability distributions over G (which randomly designate the location of the facility). If $f(\mathbf{x}) = P$, where P is a probability distribution over G , then the cost of agent i is the expected distance from x_i ,

$$\text{cost}(P, x_i) = \mathbb{E}_{y \sim P}[d(x_i, y)] .$$

A mechanism is said to be *strategyproof (SP)* if agents can never benefit by lying; formally, for every $\mathbf{x} \in G^n$, $i \in N$, and deviation $x'_i \in G$, it holds that $\text{cost}(f(x'_i, \mathbf{x}_{-i}), x_i) \geq \text{cost}(f(\mathbf{x}), x_i)$, where $\mathbf{x}_{-i} = \langle x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n \rangle$ is the vector of locations excluding x_i .

A mechanism is *group strategyproof (GSP)* if for every coalition of deviators, there is an agent that does not benefit from the deviation. Formally speaking, for every $\mathbf{x} \in G^n$, every $S \subseteq N$ and every $\mathbf{x}'_S \in G^S$, there exists $i \in S$ such that $\text{cost}(f(x'_S, \mathbf{x}_{-S}), x_i) \geq \text{cost}(f(\mathbf{x}), x_i)$. A stronger notion of group strategyproofness that is rather common in the computer science literature is obtained by requiring that for every deviation, it cannot be the case that at least one member of the coalition strictly gains while the others do not lose. While our GSP results do not hold under the stronger notion, we note that our slightly weaker notion is very common in the social choice literature, since in settings with no money (and, hence, no sidepayments) an agent has no incentive to deviate unless it strictly benefits.

We are interested in optimizing one of two target functions: the social cost, and the maximum cost. The *social cost* of a facility location $y \in G$ with respect to a preference profile \mathbf{x} is $\text{sc}(y, \mathbf{x}) = \sum_{i \in N} \text{cost}(y, x_i)$. The social cost of a distribution P with respect to \mathbf{x} is $\text{sc}(P, \mathbf{x}) = \mathbb{E}_{y \sim P}[\text{sc}(y, \mathbf{x})]$. The *maximum cost* of y with respect to \mathbf{x} is $\text{mc}(y, \mathbf{x}) = \max_{i \in N} \text{cost}(y, x_i)$. Finally, the maximum cost of a distribution P with respect to \mathbf{x} is naturally defined as $\text{mc}(P, \mathbf{x}) = \mathbb{E}_{y \sim P}[\text{mc}(y, \mathbf{x})]$.

3 Social Cost

In this section we are interested in finding a facility location that minimizes the social cost. We shall first investigate the case where the graph G is a tree; we shall then study mechanisms on general graphs, with a special emphasis on the case where the underlying graph is a circle.

3.1 Mechanisms on Trees

It is well-known that, when n agents are located on a line, the mechanism that returns the location of the median agent as the facility location is GSP. This is true since an agent can change the location of the median only by declaring itself to be on the median’s opposite side, thus pushing the median away from its true location (and this argument is easily generalized for coalitions of agents). Moreover, the median is also

optimal with respect to the social cost, since any other location with distance δ from the median is further away by δ from at least $n/2$ agents, and closer by at most δ to at most $n/2$ agents. Hence, on a line choosing the median is an optimal GSP mechanism.

If the graph G is, more generally, a tree, things are not much more complicated. Indeed, consider the following mechanism for finding the median of a tree with respect to the location profile $\mathbf{x} \in G^n$, which has been suggested in other contexts and may be considered folklore (similar ideas have been published at least as early as 1981 [6]). We first fix an arbitrary node as the root of the tree. 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.

The fact that the above mechanism is GSP is straightforward, and follows from similar arguments as the ones given for a median on a line: an agent can only modify the location of the mechanism's outcome by pushing the returned facility location away from its true location. It can also be verified that the mechanism returns a location that is optimal in terms of the social cost. To summarize, under the assumption that the graph G is a tree there is an optimal GSP mechanism for the social cost.

3.2 Mechanisms on Circles and General Graphs

We next consider the case where the graph G may contain cycles. With respect to deterministic mechanisms, the answers that we seek may be found in the paper of Schummer and Vohra. Indeed, Schummer and Vohra deal with a model that is identical to ours. They show that if G contains a cycle C , and $f : G^n \rightarrow G$ is an SP rule that is onto G , then there is a *cycle dictator*, that is, there is $i \in N$ such that for all $\mathbf{x} \in C^n$, $f(\mathbf{x}) = x_i$. In other words, if all the agents are located on C then the dictator single-handedly determines the facility location.

We claim that this result directly implies a tight SP lower bound of $n - 1$ on any graph G that contains a cycle (including a circle). Indeed, if the mechanism is not onto G then let $y \in G$ such that for all $\mathbf{x} \in G^n$, $f(\mathbf{x}) \neq y$. Now, consider the profile where all the agents are located at y ; since the optimal social cost given this profile is zero, the approximation ratio of f must be infinite. Therefore, we may assume that f is onto.

Now, if f is GSP and onto, let $C \subseteq G$ be a cycle. By the result of Schummer and Vohra, there is a cycle dictator, without loss of generality agent 1. Consider a preference profile $\mathbf{x} \in C$ where $x_1 = y \in C$ and $x_i = z \in C$ for all $i \in N \setminus \{1\}$, where $y \neq z$. We have that the optimal social cost is $d(y, z)$, by locating the facility at z , whereas $f(\mathbf{x}) = x_1 = y$, that is, $\text{sc}(f(\mathbf{x}), \mathbf{x}) = (n - 1)d(y, z)$. Hence, the approximation ratio of f is at least $n - 1$.

Enter randomization, which, as we shall demonstrate, allows us to design an SP mechanism with a constant approximation ratio. We consider the following trivial mechanism.

Mechanism 1 (Random Dictator). Given $\mathbf{x} \in G^n$, return a facility location according to the probability distribution that gives probability $1/n$ to the location x_i , for all $i \in N$.

This mechanism is obviously SP, since by deviating an agent can only lose if its own location is chosen, and does not affect the outcome if another's location is selected. We also remark that Mechanism 1 is by no means novel; it was directly employed by Meir [8] in the context of classification, and a slight variation was studied by Procaccia and Tennenholtz [10] for facility location on a line when each agent controls multiple locations. It turns out the mechanism guarantees a constant approximation ratio for the problem at hand with respect to any graph G .

Theorem 3.1. *Let G be a graph, and let $N = \{1, \dots, n\}$. Then Mechanism 1 is an SP $(2 - 2/n)$ -approximation mechanism for the social cost.*

Proof. Given $\mathbf{x} \in G^n$, let $y \in G$ be the optimal facility location, and denote $\text{OPT} = \text{sc}(y, \mathbf{x})$. Denoting Mechanism 1 by rd (for “random dictator”), it follows from the triangle inequality that

$$\begin{aligned} \text{sc}(\text{rd}(\mathbf{x}), \mathbf{x}) &= \sum_{i \in N} \frac{1}{n} \sum_{j \in N} d(x_i, x_j) \leq \frac{1}{n} \sum_{i \in N} \sum_{j \in N \setminus \{i\}} [d(x_i, y) + d(y, x_j)] \\ &= \frac{1}{n} \sum_{i \in N} [(n-1)d(x_i, y) + (\text{OPT} - d(y, x_i))] = \text{OPT} + \frac{n-2}{n} \sum_{i \in N} d(x_i, y) \\ &= \left(2 - \frac{2}{n}\right) \text{OPT} . \quad \square \end{aligned}$$

A few comments are in order. First, notice that the exact same proof actually shows that Mechanism 1 gives an approximation ratio of $2 - 2/n$ in any metric space (and, of course, the mechanism is still SP). Despite the generality of this theorem and the simplicity of its proof, we have not been able to find this explicit result in the literature, but we do not preclude the possibility that it exists in some form.

Second, it is easy to see that the analysis in the proof of Theorem 3.1 is tight. Indeed, consider a location profile where $x_1 = y \in G$ and $x_i = z \in G$ for all $j \in N \setminus \{1\}$, where $y \neq z$. Then the social cost of the optimal location is $d(y, z)$, whereas the social cost of Mechanism 1 is

$$\frac{1}{n}(n-1)d(y, z) + \frac{n-1}{n}d(z, y) = \left(2 - \frac{2}{n}\right) d(y, z) .$$

Third, in general Mechanism 1 is not GSP. To see this, let $N = \{1, 2, 3\}$ and consider a star with three arms, that is, $G = (V, E)$ where $V = \{v, u_1, u_2, u_3\}$, E contains (v, u_i) for $i = 1, 2, 3$, and $d(v, u_i) = 1$ for $i = 1, 2, 3$. In particular, $d(u_i, u_j) = 2$ for $i \neq j$. Let $\mathbf{x} \in G^n$ such that $x_i = u_i$ for $i \in N$. Then for all $i \in N$, $\text{cost}(\text{rd}(\mathbf{x}), x_i) = 4/3$, where once again we let $\text{rd}(\mathbf{x})$ denote the outcome of Mechanism 1 given \mathbf{x} . Now, consider the profile \mathbf{x}' where $x'_i = v$ for all $i \in N$, i.e., the three agents deviate to the center of the star. Since $\text{rd}(\mathbf{x}') = v$ with probability one, we have that $\text{cost}(\text{rd}(\mathbf{x}'), x_i) = 1$, hence all the agents strictly gain from the joint deviation.

On the other hand, if G is a line then Mechanism 1 is GSP. Indeed, it is quite straightforward that for any deviation, the leftmost member of the deviating coalition and the rightmost member cannot both benefit from the deviation: if the expected distance from one decreases, then the expected distance from the other increases. Much more interestingly, we have the following result.

Theorem 3.2. *Let G be a circle. Then Mechanism 1 is GSP.*

When analyzing the group strategyproofness of the mechanism, we can assume without loss of generality that the deviating coalition contains all the agents. Indeed, the expected cost of an agent given that a nondeviating agent is selected by the mechanism, and the probability that a nondeviating agent is selected by the mechanism, are both independent of the reports of the deviating agents. Hence (after scaling by a factor of n), we can give the following combinatorial, more explicit but equivalent formulation of Theorem 3.2, which may be of independent interest.

Theorem 3.3. *Let G be a circle, and let $x_1, \dots, x_n, y_1, \dots, y_n \in G$. Then there exists $i \in N$ such that*

$$\sum_{k \in N} d(x_i, x_k) \leq \sum_{k \in N} d(x_i, y_k) . \quad (1)$$

The intuition behind the proof of the theorem is as follows. When G is a line, we saw that we can identify a specific pair of agents (the leftmost and rightmost) that cannot both benefit from the deviation; this is also true when the agents are located on one semicircle, which can then be treated as an interval. However, when the agents are not on one semicircle, there are no agents that are easily identified as “extreme”. The main idea is to recognize pairs of “almost extreme” agents which we call *nearly-antipodal*. The proof establishes that there exists a pair of nearly-antipodal agents that do not both benefit from the deviation. The full proof of the theorem is relegated to Appendix A.

Summarizing and slightly generalizing the results given above, we have that if G is a line or a circle, Mechanism 1 is GSP, whereas if there is a vertex of degree at least three it is not (since any such vertex locally looks like the counterexample given above). More precisely, we have the following corollary.

Corollary 3.4. *Let G be a (connected) graph. Then Mechanism 1 is GSP if and only if the maximum degree in G is two.*

For a constant number of agents n on a circle, we were able to design a nontrivial SP mechanism with an approximation ratio slightly better than that of Mechanism 1. However, in general it is natural to ask whether Mechanism 1 is asymptotically optimal among SP approximation mechanisms for the social cost, that is: Is there an SP Mechanism whose approximation ratio for the social cost is bounded away from two? We are able to prove a lower bound of $1 + \epsilon$ for a small constant $\epsilon > 0$ (which we omit due to lack of interest), but the gap is still significant. We conjecture that the answer to our question is negative, even if G is a circle.

Conjecture 3.5. *Let $N = \{1, \dots, n\}$. Then for any circle G there is no randomized SP $(2 - \Omega(1))$ -approximation mechanism for the social cost.*

Furthermore, we conjecture that there are graphs for which it is not possible to design a GSP (randomized) mechanism with a sublinear approximation ratio.

Conjecture 3.6. *Let $N = \{1, \dots, n\}$. Then there exists a graph G such that there is no randomized GSP $o(n)$ -approximation mechanism for the social cost.*

4 Maximum Cost

In this section we shall be interested in SP mechanisms that minimize the maximum cost. The problem of designing an optimal SP mechanism is very simple with respect to deterministic mechanisms. Recall that, by the result of Schummer and Vohra [13], strategyproofness can only be obtained by a dictatorship. Hence, consider the mechanism given by $f(\mathbf{x}) = x_1$ for all $\mathbf{x} \in G^n$, that is, a dictatorship of agent 1. This mechanism is clearly GSP. Crucially, this mechanism does quite well in terms of our new optimization goal: it provides a 2-approximation for the maximum cost. To see this, given $\mathbf{x} \in G^n$, let y be the optimal facility location. Then for all $i \in N$,

$$d(x_1, x_i) \leq d(x_1, y) + d(y, x_i) \leq 2 \cdot \max\{d(y, x_1), d(y, x_i)\} \leq 2 \cdot \text{mc}(y, \mathbf{x}) .$$

On the other hand, a deterministic SP mechanism cannot achieve an approximation ratio better than two, even if the underlying graph G is a line [10]. Since in a general graph any edge is locally a line, this lower bound applies to any graph. In other words, dictatorship gives a tight SP upper bound. In the sequel we shall therefore restrict our attention to randomized mechanisms.

4.1 Randomized Mechanisms on a Circle

We presently consider the case where the graph G is a circle. An important remark is that, even using randomization, we cannot hope to achieve an SP approximation ratio better than $3/2$. Indeed, Procaccia and Tennenholtz [10] have established that a randomized SP mechanism does not yield an approximation ratio smaller than $3/2$ on a line. Furthermore, they have provided a straightforward matching GSP upper bound of $3/2$ on a line using the *Left-Right-Middle (LRM) Mechanism*: given $\mathbf{x} \in G^n$, with probability $1/4$ return the leftmost agent $\min_{i \in N} x_i$, with probability $1/4$ return the rightmost agent $\max_{i \in N} x_i$, and with probability $1/2$ return the midpoint of the interval between them, that is,

$$\text{cen} \left(\min_{i \in N} x_i, \max_{i \in N} x_i \right) = \frac{\min_{i \in N} x_i + \max_{i \in N} x_i}{2} .$$

The idea behind the strategyproofness of this mechanism is very simple: an agent can only affect the outcome of the mechanism by deviating to a location $x'_i < \min_{i \in N} x_i$ or $x'_i > \max_{i \in N} x_i$. In this case, the agent pushes the left or right boundaries *away* from its location by δ , but in doing so may push the midpoint towards its own location by $\delta/2$. Since the midpoint is selected with probability exactly twice that of each of the boundaries, the two terms cancel out.

Of course, when the agents are on a circle, in general it is meaningless to refer to the “leftmost” or “rightmost” agent. However, any semicircle can naturally be treated as an interval, and then the LRM mechanism can be applied. The mechanism that we propose is in fact a hybrid of two mechanisms: the LRM mechanism when the agents are located on one semicircle, and the *Random Center (RC) Mechanism* (defined below) when the agents are not on one semicircle.

Mechanism 2. Given $\mathbf{x} \in G^n$:

1. If \mathbf{x} is such that the agents are located on one semicircle, i.e., there exist $y, z \in G$ such that for all $i \in N$, $x_i \in [y, z]$, then we execute the LRM Mechanism on the arc $[y, z]$, treating it as an interval with the boundaries $y < z$.
2. If \mathbf{x} is such that the agents are not located on one semicircle, we execute the *Random Center (RC) Mechanism*, defined as follows.
 - (a) Randomly select a point $y \in G$.
 - (b) Let \hat{x}_i and \hat{x}_j be the two antipodal points adjacent to y , that is, \hat{x}_i is the first antipodal point encountered when walking clockwise from y , and \hat{x}_j is the first antipodal point encountered when walking counterclockwise from y .
 - (c) Return $\text{cen}(\hat{x}_i, \hat{x}_j)$.

An equivalent way of thinking about the RC Mechanism is letting the mechanism choose the center of an interval between two adjacent antipodal points with probability proportional to the length of the interval.

Some technical comments regarding Mechanism 2 are in order. Regarding the first item, there may be many choices of y and z such that $x_i \in [y, z]$ for all $i \in N$, but the LRM Mechanism is indifferent to the choice. In the context of the second item, it holds that $y \in [\hat{x}_i, \hat{x}_j]$ by the assumption that in \mathbf{x} the agents are not on one semicircle, i.e., y is on the same arc whose center we return. Furthermore, the RC mechanism is ambiguously defined when the random point y is an antipodal point itself, but this happens with probability zero.

A priori, it seems there is no reason to suppose that Mechanism 2 is SP, as usually such hybridizations of SP mechanisms are not SP as a whole. Nevertheless, we have the following theorem.

Theorem 4.1. *Assume that G is a circle. Then Mechanism 2 is an SP 3/2-approximation mechanism for the maximum cost.*

The nontrivial part of the theorem’s proof is the strategyproofness of the mechanism. Although the proof is long, it revolves around several basic properties of the RC Mechanism. Very generally speaking, one of the ideas at the core of the proof is that the locations occupied by the agents in \mathbf{x} are special, but only in the sense that for every $i \in N$ the antipodal point \hat{x}_i is among the antipodal points. For this reason, the cost of an agent (assuming that the circumference of the circle is one) under the RC Mechanism is at most 1/4 (Lemma B.5). A second important idea is that, from the point of view of agent i , the RC Mechanism essentially chooses a location uniformly on the circle, except for its behavior on the arc between the two antipodal points adjacent to x_i (Lemma B.4). The detailed proof is given in Appendix B.

It is possible to show that Mechanism 2 is not GSP, even when the agents are assumed not to be located on one semicircle before and after the deviation (that is, the RC Mechanism is used in both cases). The counterexample is due to Dror Shemesh.

4.2 Randomized Mechanisms on Trees

In the following we assume that the graph G is a tree. We first observe that randomization allows us to do slightly better than dictatorship, especially when the number of agents is small. Indeed, given $\mathbf{x} \in G^n$, the center of G with respect to \mathbf{x} is a point

$$y \in \operatorname{argmin}_{z \in G} \operatorname{mc}(z, \mathbf{x}) .$$

It is easy to verify that when G is a tree the center is unique.² Therefore, we can denote the (unique) center of G with respect to \mathbf{x} by $\operatorname{cen}(G, \mathbf{x})$.

We consider the following mechanism: given $\mathbf{x} \in G^N$, the distribution on the returned location gives probability $1/(n+2)$ to x_i for each $i \in N$, and probability $2/(n+2)$ to $\operatorname{cen}(G, \mathbf{x})$. The fact that the mechanism is SP follows from the fact that when agent i deviates from x_i to x'_i , it holds that

$$|d(x_i, \operatorname{cen}(G, \mathbf{x})) - d(x_i, \operatorname{cen}(G, \mathbf{x}'))| \leq \frac{d(x_i, x'_i)}{2} ,$$

and therefore, denoting the above mechanism by f ,

$$\operatorname{cost}(f(\mathbf{x}'), x_i) - \operatorname{cost}(f(\mathbf{x}), x_i) \geq \frac{1}{n} \cdot d(x_i, x'_i) - \frac{2}{n} \cdot \frac{d(x_i, x'_i)}{2} = 0 .$$

The approximation ratio of the mechanism satisfies

$$\frac{\operatorname{mc}(f(\mathbf{x}), \mathbf{x})}{\operatorname{OPT}} \leq \frac{\frac{2}{n+2} \cdot \operatorname{OPT} + \frac{n}{n+2} \cdot 2 \cdot \operatorname{OPT}}{\operatorname{OPT}} = 2 - \frac{2}{n+2} .$$

Despite this small improvement over dictatorship, we shall demonstrate that we cannot do significantly better. In other words, our final major result asserts that an SP mechanism cannot achieve an approximation ratio that is bounded away from two for the maximum cost, even on trees.

Theorem 4.2. *Let $N = \{1, \dots, n\}$. Then there exists a tree G such that no SP randomized mechanism can have an approximation ratio that is smaller than $2 - \mathcal{O}\left(\frac{1}{2^{\sqrt{\log n}}}\right)$ for the maximum cost.*

The proof of the theorem is given in Appendix C.

²This would not be true in a discrete graph model, but this issue can still be easily circumvented.

References

- [1] S. Arora, P. Raghavan, and S. Rao. Approximation schemes for Euclidean k -medians and related problems. In *Proceedings of the 30th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 106–113, 1998.
- [2] M. Bern and D. Eppstein. Approximation algorithms for geometric problems. In D. Hochbaum, editor, *Approximation Algorithms for NP-Hard Problems*. PWS Publishing, 1996.
- [3] H. J. Brandelt and M. Labbé. How bad can a voting location be. *Social Choice and Welfare*, 3:125–145, 1986.
- [4] O. Dekel, F. Fischer, and A. D. Procaccia. Incentive compatible regression learning. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 277–286, 2008.
- [5] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–602, 1973.
- [6] P. Hansen and J.-F. Thisse. Outcomes of voting and planning: Condorcet, Weber and Rawls locations. *Journal of Public Economics*, 16:1–15, 1981.
- [7] M. Labbé. Outcomes of voting and planning in single facility location problems. *European Journal of Operations Research*, 20:299–313, 1985.
- [8] R. Meir. Strategy proof classification. Master’s thesis, The Hebrew University of Jerusalem, 2008. Available from: <http://www.cs.huji.ac.il/~reshef24/spc.thesis.pdf>.
- [9] N. Nisan. Introduction to mechanism design (for computer scientists). In N. Nisan, T. Roughgarden, É. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*, chapter 9. Cambridge University Press, 2007.
- [10] A. D. Procaccia and M. Tennenholtz. Approximate mechanism design without money. In *Proceedings of the 10th ACM Conference on Electronic Commerce (ACM-EC)*, 2009. To appear.
- [11] M. Rothkopf. Thirteen reasons the Vickrey-Clarke-Groves process is not practical. *Operations Research*, 55(2):191–197, 2007.
- [12] M. Satterthwaite. Strategy-proofness and Arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.
- [13] J. Schummer and R. V. Vohra. Strategy-proof location on a network. *Journal of Economic Theory*, 104(2):405–428, 2004.
- [14] J. Schummer and R. V. Vohra. Mechanism design without money. In N. Nisan, T. Roughgarden, É. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*, chapter 10. Cambridge University Press, 2007.

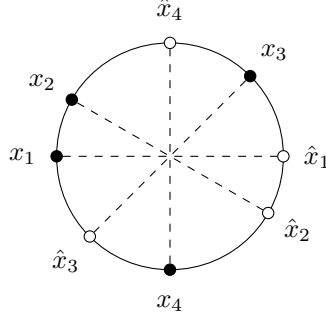


Figure 1: An illustration of the construction in the proof of Theorem 3.3, for $n = 4$. The nearly antipodal pairs are $A = \{\langle x_1, x_3 \rangle, \langle x_2, x_4 \rangle, \langle x_3, x_4 \rangle\}$.

A Proof of Theorem 3.3

Assume without loss of generality that the circumference of G is 1. Let $x_1, \dots, x_n \in G$, and define a multiset X by $X = \{x_1, \dots, x_n\}$. We first note that we can assume that there are no $x_i, x_j \in X$ such that $x_j = \hat{x}_i$. Indeed, in this case the claim holds trivially with respect to either i or j , since for all $z \in G$,

$$d(x_i, z) + d(x_j, z) = 1/2 .$$

In particular, for every $x_i, x_j \in X$, (x_i, x_j) and (\hat{x}_i, \hat{x}_j) are well-defined.

We say that two points $x_i, x_j \in X$ are *nearly antipodal* if there is no point $x_k \in X$ such that $x_k \in (x_i, \hat{x}_j)$ or $x_k \in (x_j, \hat{x}_i)$; let $A \subseteq X^2$ be the set of all nearly antipodal pairs. Given a nearly antipodal pair $\langle x_i, x_j \rangle \in A$, let the *critical arc of $\langle x_i, x_j \rangle$* , denoted $\text{crit}(x_i, x_j)$, be the *long* open arc between \hat{x}_i and \hat{x}_j , that is,

$$\text{crit}(x_i, x_j) = G \setminus [\hat{x}_i, \hat{x}_j] = (x_i, \hat{x}_j) \cup [x_i, x_j] \cup (x_j, \hat{x}_i) .$$

See Figure 1 for an illustration of the construction given above.

Let $y_1, \dots, y_n \in G$, and define a multiset Y by $Y = \{y_1, \dots, y_n\}$. It is sufficient to prove that there exists a pair of nearly antipodal points $\langle x_i, x_j \rangle \in A$ such that

$$\sum_{k \in N} d(x_i, x_k) + \sum_{k \in N} d(x_j, x_k) \leq \sum_{k \in N} d(x_i, y_k) + \sum_{k \in N} d(x_j, y_k) .$$

Indeed, in this case we get that Equation (1) holds with respect to either x_i or x_j . Therefore, assume for the purpose of contradiction that for every pair of nearly antipodal points $\langle x_i, x_j \rangle \in A$,

$$\sum_{k \in N} d(x_i, x_k) + \sum_{k \in N} d(x_j, x_k) > \sum_{k \in N} d(x_i, y_k) + \sum_{k \in N} d(x_j, y_k) . \quad (2)$$

We claim that Equation (2) implies that for every pair of nearly antipodal points $\langle x_i, x_j \rangle \in A$, the number of points from Y on $\text{crit}(x_i, x_j)$ is strictly greater than the number of points from X on the same arc. Formally, for $\langle x_i, x_j \rangle \in A$, let

$$\alpha_{ij}^X = |\{x_k \in X : x_k \in \text{crit}(x_i, x_j)\}| ,$$

and

$$\alpha_{ij}^Y = |\{y_k \in Y : y_k \in \text{crit}(x_i, x_j)\}| .$$

We have the following claim.

Lemma A.1. *Let $\langle x_i, x_j \rangle \in A$. Then $\alpha_{ij}^Y > \alpha_{ij}^X$.*

Proof. For every point $z \in [x_i, x_j]$, we have that

$$d(x_i, z) + d(x_j, z) = d(x_i, x_j) .$$

Let $d'(x_i, x_j)$ be the length of the longer arc $G \setminus [x_i, x_j]$ between x_i and x_j , namely

$$d'(x_i, x_j) = d(x_i, \hat{x}_j) + d(\hat{x}_j, \hat{x}_i) + d(\hat{x}_i, x_j) = d(x_i, x_j) + 2 \cdot d(x_i, \hat{x}_j) > d(x_i, x_j) .$$

For every point $z \in [\hat{x}_i, \hat{x}_j]$ it holds that

$$d(x_i, z) + d(x_j, z) = d'(x_i, x_j) .$$

Finally, it holds that for every $z \in (x_i, \hat{x}_j) \cup (x_j, \hat{x}_i)$,

$$d(x_i, x_j) < d(x_i, z) + d(x_j, z) < d'(x_i, x_j) .$$

Since x_i and x_j are nearly antipodal, there are no points from X in (x_i, \hat{x}_j) and (x_j, \hat{x}_i) . Therefore,

$$\sum_{k \in N} d(x_i, x_k) + \sum_{k \in N} d(x_j, x_k) = \alpha_{ij}^X \cdot d(x_i, x_j) + (n - \alpha_{ij}^X) \cdot d'(x_i, x_j) . \quad (3)$$

On the other hand,

$$\sum_{k \in N} d(x_i, y_k) + \sum_{k \in N} d(x_j, y_k) \geq \alpha_{ij}^Y \cdot d(x_i, x_j) + (n - \alpha_{ij}^Y) \cdot d'(x_i, x_j) . \quad (4)$$

Using Equations (3) and (4), we get that (2) directly implies that $\alpha_{ij}^Y > \alpha_{ij}^X$, as claimed. \square

From Lemma A.1, we immediately get that

$$\sum_{(x_i, x_j) \in A} \alpha_{ij}^X < \sum_{(x_i, x_j) \in A} \alpha_{ij}^Y . \quad (5)$$

In order to derive a contradiction, we also need the following lemma.

Lemma A.2. *There exists $r \in \mathbb{N}$ such that*

$$\sum_{(x_i, x_j) \in A} \alpha_{ij}^X = r \cdot n , \quad (6)$$

whereas

$$\sum_{(x_i, x_j) \in A} \alpha_{ij}^Y \leq r \cdot n , \quad (7)$$

Proof. It is easy to see that $|A|$ is odd (e.g., by induction on n); let $|A| = 2s + 1$, for some $s \in \mathbb{N}$. We first wish to claim that every $x_i \in X$ is a member of exactly $s + 1$ critical arcs, which directly proves Equation (6) with $r = s + 1$.

Without loss of generality we prove the claim with respect to $x_1 \in X$. Consider the clockwise closed arc between x_1 and \hat{x}_1 . Let $Z = \{z_1, \dots, z_t\}$ be all the points x_i or \hat{x}_i on this arc, where for all k , $z_{k+1} \succeq z_k$. In particular, $z_1 = x_1$ and $z_t = \hat{x}_1$. For instance, in Figure 1 we have that $Z = \{x_1, x_2, \hat{x}_4, x_3, \hat{x}_1\}$.

Now, we have that the set of nearly-antipodal pairs A is exactly the set of pairs $\langle x_i, x_j \rangle$ such that z_k is a point x_i and z_{k+1} is an antipodal point \hat{x}_j (this is a *type 1* nearly-antipodal pair), or z_k is an antipodal point \hat{x}_i and z_{k+1} is a point x_j (this is a *type 2* nearly-antipodal pair). If $\langle x_i, x_j \rangle$ is a nearly-antipodal pair of type 1, we have that $x_1 \in [x_i, x_j]$, and hence $x_1 \in \text{crit}(x_i, x_j)$. On the other hand, if $\langle x_i, x_j \rangle$ is a nearly-antipodal pair of type 2, then $x_1 \notin \text{crit}(x_i, x_j)$. Since $z_1 = x_1$ is a point from X and $z_{n+1} = \hat{x}_1$ is an antipodal point, the number of nearly-antipodal pairs of type 1 is exactly $s + 1$, which proves the claim.

In order to prove Equation (7), let $y \in G$. It is sufficient to prove that there exists $x_i \in X$ such that y appears in at most as many critical arcs as x_i , since we already know that x_i is a member of exactly $s + 1$ critical arcs. We consider the two points or antipodal points that are adjacent to y , and briefly examine four cases.

1. $x_i \preceq y \preceq x_j$: y appears in exactly the critical arcs that contain x_i (these are also exactly the critical arcs that contain x_j).
2. $x_i \preceq y \prec \hat{x}_j$: y appears in exactly the critical arcs that contain x_i .
3. $\hat{x}_i \prec y \preceq x_j$: y appears in exactly the critical arcs that contain x_j .
4. $\hat{x}_i \preceq y \preceq \hat{x}_j$: When walking counterclockwise from \hat{x}_i , let $x_k \in X$ be the first point from X , and let \hat{x}_l be the last antipodal point such that $x_k \prec \hat{x}_l \preceq \hat{x}_i$. Then y is contained in all the critical arcs that contain x_k , except for $\text{crit}(x_k, \hat{x}_l)$, that is, in exactly s critical arcs.

We deduce that every y_i is contained in at most $r = s + 1$ critical arcs, which implies the validity of Equation (7). \square

It follows from Lemma A.2 that

$$\sum_{(x_i, x_j) \in A} \alpha_{ij}^X \geq \sum_{(x_i, x_j) \in A} \alpha_{ij}^Y,$$

in contradiction to Equation (5). \square

B Proof of Theorem 4.1

In the proof we assume without loss of generality that the circumference of the circle G is 1. In addition, we denote the outcome of the LRM Mechanism and the RC Mechanism given $\mathbf{x} \in G^n$ by $\text{lrm}(\mathbf{x})$ and $\text{rc}(\mathbf{x})$, respectively.

Lemma B.1. *Mechanism 2 is a 3/2-approximation mechanism for the maximum cost.*

Proof. Assume first that $\mathbf{x} \in G^n$ is such that not all agents are located on one semicircle. Let α be the length of the longest arc between two adjacent agents, and assume without loss of generality that these agents are agents 1 and 2. Since the agents are not located on one semicircle, it holds that $\alpha \leq 1/2$. It can be verified

that the optimal facility location is $\text{cen}(\hat{x}_1, \hat{x}_2)$, hence we have that $\text{OPT} = (1 - \alpha)/2$. The mechanism selects the optimal solution with probability α , and with probability $1 - \alpha$ selects a solution with maximum cost at most $1/2$. Therefore, the approximation ratio is at most

$$\frac{\text{mc}(\text{rc}(\mathbf{x}), \mathbf{x})}{\text{OPT}} \leq \frac{\alpha \cdot \frac{1-\alpha}{2} + (1-\alpha) \cdot \frac{1}{2}}{\frac{1-\alpha}{2}} = 1 + \alpha \leq \frac{3}{2}. \quad (8)$$

If \mathbf{x} is such that all the agents are located on one semicircle, then the LRM mechanism is applied: we choose the optimal location with probability $1/2$, and a location with twice the optimal cost with probability $1/2$, hence the approximation ration is, once again, $3/2$. \square

In order to establish the strategyproofness of Mechanism 2, we must examine four types of lies: an agent deviating such that before the deviation all the agents were located on one semicircle and after the deviation they are located on one semicircle—“semicircle to semicircle” (the LRM Mechanism is applied to both); not semicircle to semicircle (RC to LRM); semicircle to not semicircle (LRM to RC); and not semicircle to not semicircle (RC to RC). The semicircle to semicircle case is relatively straightforward, and we tackle it first.

Lemma B.2 (Semicircle to semicircle). *Assume that $\mathbf{x} \in G^n$ is such that the agents are on one semicircle, and agent i deviates such that in the new location profile \mathbf{x}' the agents are also on one semicircle. Then*

$$\text{cost}(\text{lrn}(\mathbf{x}), x_i) \leq \text{cost}(\text{lrn}(\mathbf{x}'), x_i) .$$

Proof. Let $x_1 \succeq x_2 \succeq \dots \succeq x_n$, and denote $x_1 = l$ (for “left”) and $x_n = r$ (for “right”). Suppose agent i deviates from x_i to x'_i , such that the agents are on one semicircle. If (x'_i, x_i) intersects with the new semicircle, the proof follows directly from the fact that the LRM Mechanism is SP when applied to an interval [10].

It is easy to verify that (x'_i, x_i) may lie in the complement of the new semicircle only if (i) the deviating agent is r and $x'_i \succ l$; or (ii) the deviating agent is l and $x'_i \prec r$. We prove the lemma for the former case, but note that the latter case is completely analogous.

Indeed, suppose without loss of generality that $x_i = r$ and $x'_i \succ l$. Let r' be the adjacent agent to r such that $r' \succ r$. It must hold that $x'_i \preceq r'$ (otherwise in the new location profile the agents are not all on one semicircle). We denote $\alpha = d(r, r')$, $\beta = d(x'_i, l)$, and $\gamma = d(l, r)$ (see Figure 2). From the assumptions of the lemma it follows that $\gamma \leq \frac{1}{2}$. The short arc from x'_i to x_i lies in the complement of the new semicircle if and only if $1 - \beta - \gamma \leq \beta + \gamma$, that is, $\beta + \gamma \geq 1/2$.

We first calculate the cost of x_i in the location profile \mathbf{x} :

$$\text{cost}(\text{lrn}(\mathbf{x}), x_i) = \frac{1}{4} \cdot \gamma + \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot \frac{\gamma}{2} = \frac{\gamma}{2} .$$

We distinguish between two cases:

Case 1: $\alpha \leq 1 - \beta - \gamma$. In this case, the cost of the LRM Mechanism in the new location profile with respect to x_i is

$$\text{cost}(\text{lrn}(\mathbf{x}'), x_i) = \frac{1}{4} \cdot \alpha + \frac{1}{4} \cdot (1 - \beta - \gamma) + \frac{1}{2} \cdot \left(\alpha + \frac{\gamma - \alpha + \beta}{2} \right) = \frac{1}{4} + \frac{\alpha}{2} .$$

It holds that

$$\text{cost}(\text{lrn}(\mathbf{x}), x_i) \leq \text{cost}(\text{lrn}(\mathbf{x}'), x_i) \Leftrightarrow \gamma \leq \frac{1}{2} + \alpha ,$$

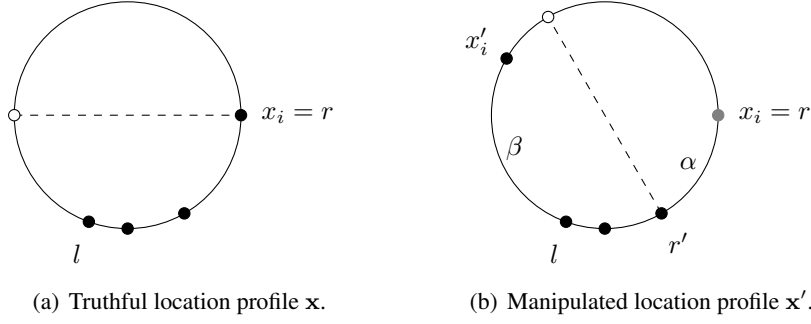


Figure 2: Illustration of the proof of Lemma B.2.

but this holds since $\gamma \leq \frac{1}{2}$.

Case 2: $\alpha > 1 - \beta - \gamma$. In this case, the cost of the left-middle-right mechanism in the new location profile with respect to x_i is

$$\text{cost}(\text{lrm}(\mathbf{x}'), x_i) = \frac{1}{4} \cdot \alpha + \frac{1}{4} \cdot (1 - \beta - \gamma) + \frac{1}{2} \cdot \left(1 - \beta - \gamma + \frac{\gamma - \alpha + \beta}{2} \right) = \frac{3}{4} - \frac{\beta}{2} - \frac{\gamma}{2} .$$

It holds that

$$\text{cost}(\text{lrm}(\mathbf{x}), x_i) \leq \text{cost}(\text{lrm}(\mathbf{x}'), x_i) \Leftrightarrow \frac{\gamma}{2} \leq \frac{3}{4} - \frac{\beta}{2} - \frac{\gamma}{2} \Leftrightarrow \gamma \leq \frac{3}{4} - \frac{\beta}{2} .$$

Since $\gamma \leq \frac{1}{2}$, it is sufficient to show that $\frac{1}{2} \leq \frac{3}{4} - \frac{\beta}{2}$, which holds if and only if $\beta \leq \frac{1}{2}$. The last inequality follows from the fact that all the agents are on one semicircle after the deviation; this concludes the proof of the lemma. \square

The other three deviations are more complicated, and their proofs require laying some foundations first. We start with two simple lemmata concerning partitions of intervals on the line, both of which will prove useful in several points in the sequel. On a circle, the arc between two points $x, y \in G$ such that $d(x, y) \leq 1/2$ can be regarded as an interval.

Lemma B.3. *Let $x, y \in [0, 1] \subset \mathbb{R}$. It holds that*

$$d(x, 1/2) \leq y \cdot d(x, \text{cen}(0, y)) + (1 - y) \cdot d(x, \text{cen}(y, 1)) .$$

Proof. Assume without loss of generality that $x \leq 1/2$. We distinguish between two cases.

Case 1: $y/2 \geq x$. In this case,

$$y \cdot d(x, \text{cen}(0, y)) + (1 - y) \cdot d(x, \text{cen}(y, 1)) = y \cdot \left(\frac{y}{2} - x \right) + (1 - y) \cdot \left(y + \frac{1 - y}{2} - x \right) = \frac{1}{2} - x .$$

Case 2: $y/2 < x$. We have that

$$y \cdot d(x, \text{cen}(0, y)) + (1 - y) \cdot d(x, \text{cen}(y, 1)) = y \cdot \left(x - \frac{y}{2} \right) + (1 - y) \cdot \left(y + \frac{1 - y}{2} - x \right) = \frac{1}{2} - x + (2xy - y^2) .$$

Since $y/2 < x$, it holds that $2xy > y^2$, hence $1/2 - x + (2xy - y^2) > 1/2 - x$. \square

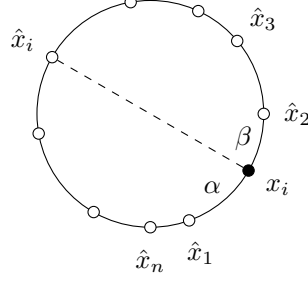


Figure 3: Illustration of the proof of Lemma B.5.

Lemma B.4. Let $y_1, \dots, y_{m+1} \in [0, 1/2]$ such that $y_1 = 0, y_{m+1} = 1/2$. For all $i = 1, \dots, m$, define $d_i = y_{i+1} - y_i$. Then

$$\sum_{i=1}^m \left(d_i \cdot \left(\sum_{j=1}^{i-1} d_j + \frac{d_i}{2} \right) \right) = \frac{1}{8} .$$

The intuition behind Lemma B.4 is that choosing the center of each interval with probability equal to the length of the interval is like randomly choosing a point in $[0, 1/2]$ with probability $1/2$. The expected distance of a random point in $[0, 1/2]$ from 0 is $1/4$, multiplying by $1/2$ yields $1/8$.

Proof of Lemma B.4. By reorganizing the terms, it can be verified that

$$\sum_{i=1}^m \left(d_i \cdot \left(\sum_{j=1}^{i-1} d_j + \frac{d_i}{2} \right) \right) = \frac{(\sum_{i=1}^m d_i)^2}{2} = \frac{1}{8} .$$

□

The next lemma implies that under the Random Center Mechanism, the cost of an agent is at most $1/4$. It follows from the two previous lemmata.

Lemma B.5. For all $\mathbf{x} \in G^m$ such that \mathbf{x} is not on one semicircle, and for all $i \in N$,

$$\text{cost}(\text{rc}(\mathbf{x}), x_i) \leq \frac{1}{4} .$$

Proof. Assume that the locations of the agents satisfy $\hat{x}_1 \succ \hat{x}_2 \cdots \succ \hat{x}_n \succ \hat{x}_1$, and that there are no antipodal points in $(\hat{x}_i, \hat{x}_{i+1})$ for all $i \in N$. Let $i \in N$, and assume without loss of generality that $x_i \in [\hat{x}_1, \hat{x}_2]$. In addition, let $\alpha = d(\hat{x}_1, x_i)$ and $\beta = d(x_i, \hat{x}_2)$ (see Figure B).

We wish to calculate the cost of the Random Center Mechanism with respect to agent i . We can break this cost down into two components: the cost when choosing a point in the arc $[\hat{x}_1, \hat{x}_2]$ (this happens with probability $\alpha + \beta$), and the rest of the cost. Notice that $x_i \in [\hat{x}_1, \hat{x}_2]$, i.e., the length of the arc between \hat{x}_1 and \hat{x}_2 that includes x_i is at most $1/2$; this holds by the assumption that the agents are not on one semicircle. Therefore we can treat $[\hat{x}_1, \hat{x}_2]$ as an interval. By normalizing $d(\hat{x}_1, \hat{x}_2)$, we get from Lemma B.3 that if we partition the arc $[\hat{x}_1, \hat{x}_2]$ into two arcs of length y and $1 - y$ and choose the center of each with probability y and $1 - y$, respectively, the cost of agent i can only increase. In particular, this is true when $y = \alpha$. Hence, we have that

$$\text{cost}(\text{rc}(\mathbf{x}), x_i) \leq \text{cost}(\text{rc}(\mathbf{x} \cup \{\hat{x}_i\}), x_i) , \quad (9)$$

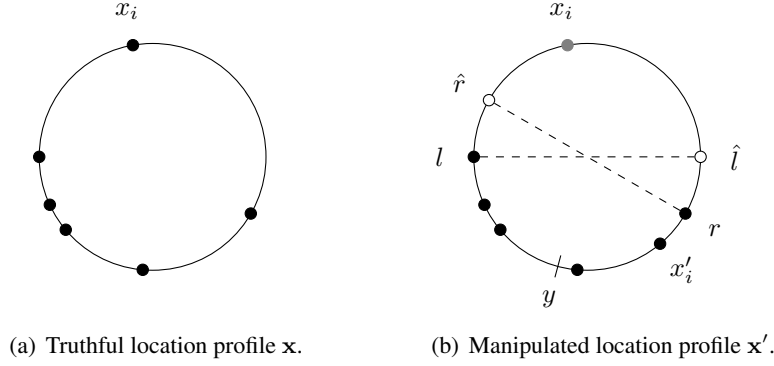


Figure 4: Illustration of the proof of Lemma B.6.

where $\mathbf{x} \cup \{\hat{x}_i\}$ is the profile \mathbf{x} with an additional agent at \hat{x}_i , that is, an additional antipodal point at x_i . It is sufficient to show that

$$\text{cost}(\text{rc}(\mathbf{x} \cup \{\hat{x}_i\}), x_i) \leq \frac{1}{4} .$$

The expression $\text{cost}(\text{rc}(\mathbf{x} \cup \{\hat{x}_i\}), x_i)$ is the expected distance from agent i when the center of one of the arcs

$$[x_i, \hat{x}_2], [\hat{x}_2, \hat{x}_3], [\hat{x}_3, \hat{x}_4], \dots, [\hat{x}_n, \hat{x}_1], [\hat{x}_1, x_i] , \quad (10)$$

is chosen, where the probability of choosing the center of an arc is its length. In order to make this explicit, denote $d_i = d(\hat{x}_i, \hat{x}_{i+1})$ for $i = 1, \dots, n-1$, $d_n = d(\hat{x}_n, \hat{x}_1)$. We have

$$\text{cost}(\text{rc}(\mathbf{x} \cup \{\hat{x}_i\}), x_i) = \frac{\alpha^2}{2} + \frac{\beta^2}{2} + \sum_{j=2}^{i-1} \left(d_j \cdot \left(\beta + \sum_{k=2}^{j-1} d_k + \frac{d_j}{2} \right) \right) + \sum_{j=i}^n \left(d_j \cdot \left(\alpha + \sum_{k=j+1}^n d_k + \frac{d_j}{2} \right) \right) \quad (11)$$

We partition the expression in the right hand side of Equation (11) into two sums, each corresponding to the cost of the mechanism on a semicircle of length $1/2$, and apply Lemma B.4 to each, with $y_1 = x_i$ in both cases, and $d_1 = \alpha$ or $d_1 = \beta$. In more detail, it holds that

$$\frac{\beta^2}{2} + \sum_{j=2}^{i-1} \left(d_j \cdot \left(\beta + \sum_{k=2}^{j-1} d_k + \frac{d_j}{2} \right) \right) = \frac{1}{8} ,$$

and

$$\frac{\alpha^2}{2} + \sum_{j=i}^n \left(d_j \cdot \left(\alpha + \sum_{k=j+1}^n d_k + \frac{d_j}{2} \right) \right) = \frac{1}{8} .$$

We conclude that the expression on the right hand side of Equation (11) is exactly $1/4$. \square

Lemma B.6 (Not semicircle to semicircle). *Assume that $\mathbf{x} \in G^n$ is such that the agents are not on one semicircle, but agent i deviates such that in the new location profile \mathbf{x}' the agents are on one semicircle. Then*

$$\text{cost}(\text{rc}(\mathbf{x}), x_i) \leq \text{cost}(\text{lm}(\mathbf{x}'), x_i) .$$

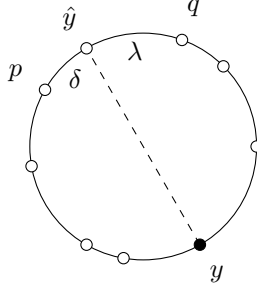


Figure 5: Illustration of the proof of Lemma B.7.

Proof. By Lemma B.5 we have that $\text{cost}(\text{rc}(\mathbf{x}), x_i) \leq 1/4$, therefore it will be sufficient to prove that $\text{cost}(\text{lrm}(\mathbf{x}'), x_i) \geq 1/4$.

Let l (for “left”) and r for “right” be the two extreme agent locations in \mathbf{x}' , where $l \succeq r$. Note that since the agents were not on one semicircle under \mathbf{x} , we have that $x_i \in (\hat{l}, \hat{r})$. Let $y = \text{cen}(l, r)$ be the center of the arc (l, r) (see Figure 4). We claim that $d(x_i, y) \geq 1/4$. This follows immediately from the facts that $d(\hat{l}, y) \geq 1/4$, $d(\hat{r}, y) \geq 1/4$, and $x_i \in (\hat{l}, \hat{r})$. Hence, the cost of the mechanism is at least

$$\text{cost}(\text{lrm}(\mathbf{x}'), x_i) = \frac{1}{4} \cdot d(x_i, l) + \frac{1}{4} \cdot d(x_i, r) + \frac{1}{2} \cdot d(x_i, y) \geq \frac{1}{4} \cdot (d(x_i, l) + d(x_i, r)) + \frac{1}{2} \cdot \frac{1}{4} \geq \frac{1}{4} ,$$

where the last transition follows from the fact that $d(x_i, l) + d(x_i, r)$ is the length of the *long* arc between l and r , therefore the value of this sum is at least $1/2$. \square

In order to deal with the last two deviations, we require one additional fundamental lemma. The lemma asserts that the cost of the RC Mechanism with respect to a point y can only decrease if the point y is added to the vector of locations. This is, in fact, the mirror image of Equation (9), which is itself a special case of Lemma B.3.

Lemma B.7. *Let $\mathbf{x} \in G^n$ such that \mathbf{x} is not on one semicircle, and let $y \in G$. Then*

$$\text{cost}(\text{rc}(\mathbf{x}), y) \geq \text{cost}(\text{rc}(\mathbf{x} \cup \{y\}), y) .$$

Proof. The cost incurred from the mechanism $\text{rc}(\mathbf{x})$ with respect to y is identical to the cost incurred from $\text{rc}(\mathbf{x} \cup \{y\})$ for all the intervals, except for the interval the point \hat{y} is on. Let p and q denote the antipodal points adjacent to \hat{y} such that $p \preceq \hat{y} \preceq q$. Denote $\delta = d(p, \hat{y})$ and $\lambda = d(\hat{y}, q)$, and assume without loss of generality that $\delta \leq \lambda$ (see Figure 5).

It is sufficient to show that the cost incurred when the random chosen point is on the arc $[p, q]$ is lower under $\text{rc}(\mathbf{x} \cup \{y\})$ than under $\text{rc}(\mathbf{x})$. The cost incurred by points on the arc $[p, q]$ under $\text{rc}(\mathbf{x} \cup \{y\})$ is $\delta(d(y, p) + \delta/2) + \lambda(d(y, q) + \lambda/2)$. The cost incurred by points on the arc $[p, q]$ under $\text{rc}(\mathbf{x})$ is $(\delta + \lambda)(d(y, q) + \frac{\delta + \lambda}{2})$. It holds that

$$(\delta + \lambda) \left(d(y, q) + \frac{\delta + \lambda}{2} \right) \geq \delta \left(d(y, p) + \frac{\delta}{2} \right) + \lambda \left(d(y, q) + \frac{\lambda}{2} \right) \Leftrightarrow d(y, q) + \lambda \geq d(y, p).$$

The inequality on the right hand side holds since $d(y, q) + \lambda = \frac{1}{2}$ while $d(y, p) \leq \frac{1}{2}$. \square

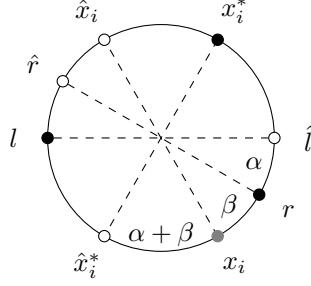


Figure 6: Illustration of the proof of Lemma B.8.

Lemma B.8 (Semicircle to not semicircle). *Assume that $\mathbf{x} \in G^n$ is such that the agents are on one semicircle, but agent i deviates such that in the new location profile \mathbf{x}' the agents are not on one semicircle. Then*

$$\text{cost}(\text{lrn}(\mathbf{x}), x_i) \leq \text{cost}(\text{rc}(\mathbf{x}'), x_i) .$$

Proof. Let $x_1 \succeq x_2 \cdots \succeq x_n$ be the location of the n agents, and denote $l = x_1$ (the leftmost agent) and $r = x_n$. Let $i \in N$, and assume without loss of generality that $d(x_i, r) \leq d(x_i, l)$, or equivalently $d(x_i, \hat{l}) \leq d(x_i, \hat{r})$. Let $\alpha = d(r, \hat{l})$, and $\beta = d(r, x_i)$. Finally, let x'_i denote the new location of agent i (see Figure 6).

We first calculate the cost of x_i in the location profile \mathbf{x} under the LRM mechanism. We have that $d(x_i, l) = \frac{1}{2} - \alpha - \beta$, $d(r, l) = \frac{1}{2} - \alpha$, and

$$d(x_i, \text{cen}(l, r)) = \frac{1}{4} - \frac{\alpha}{2} - \beta .$$

Then

$$\text{cost}(\text{lrn}(\mathbf{x}), x_i) = \frac{\beta}{4} + \frac{\frac{1}{2} - \alpha - \beta}{4} + \frac{\frac{1}{4} - \frac{\alpha}{2} - \beta}{2} = \frac{1}{4} - \frac{\alpha}{2} - \frac{\beta}{2} . \quad (12)$$

We now wish to give a lower bound on $\text{cost}(\text{rc}(\mathbf{x}'), x_i)$. First, by Lemma B.7 we have that

$$\text{cost}(\text{rc}(\mathbf{x}'), x_i) \geq \text{cost}(\text{rc}(\mathbf{x}' \cup \{x_i\}), x_i) .$$

A subtle remark is that the above argument allows us to take \hat{r} into account even if $x_i = r$ and agent i deviated to x'_i , a fact that allows us to avoid distinguishing this extreme case.

Next, we would like to fix an “optimal” location for \hat{x}'_i . First notice that since \mathbf{x}' is not on one semicircle, it must hold that $x'_i \in (\hat{r}, \hat{l})$, that is, $\hat{x}'_i \in (r, l)$. Now, we ask: what is the location of \hat{x}'_i that yields the lowest cost with respect to x_i , when the RC Mechanism chooses a point in $G \setminus (\hat{l}, \hat{r})$ (the long arc between \hat{l} and \hat{r})? There are no antipodal points, other than \hat{x}'_i , in $G \setminus (\hat{l}, \hat{r})$. Furthermore, we can treat this arc as an interval with respect to distances from x_i , since the short arcs $[x_i, \hat{l}]$ and $[x_i, \hat{r}]$ are contained in $G \setminus (\hat{l}, \hat{r})$. Therefore, we can apply Lemma B.3 (by normalizing the length of $G \setminus (\hat{l}, \hat{r})$). In particular, let $\hat{x}^*_i \succeq x_i$ such that

$$d(\hat{x}^*_i, x_i) = d(x_i, \hat{l}) = \alpha + \beta ,$$

and let $\mathbf{x}^* = \langle x^*_i, x_{-i} \rangle$; see Figure 6 for an illustration. From (the proof of) Lemma B.3 it follows that

$$\text{cost}(\text{rc}(\mathbf{x}' \cup \{x_i\}), x_i) \geq \text{cost}(\text{rc}(\mathbf{x}^* \cup \{x_i\}), x_i) .$$

Hence, it is sufficient to prove that

$$\text{cost}(\text{rc}(\mathbf{x}^* \cup \{x_i\}), x_i) \geq \frac{1}{4} - \frac{\alpha}{2} - \frac{\beta}{2} . \quad (13)$$

We calculate the cost of the mechanism $\text{rc}(\mathbf{x}^* \cup \{x_i\})$ with respect to agent i . Let y be the point chosen by the mechanism. By the choice of x_i^* , if y is on the arc $[\hat{l}, \hat{x}_i^*]$, the cost is zero. If $y \in [\hat{l}, \hat{x}_i]$, we apply Lemma B.4 with $y_1 = x_i, y_{m+1} = \hat{x}_i, y_j = \hat{x}_{j-1}$ for $j = 2, \dots, m+1$, and $d_1 = \alpha + \beta$. We get:

$$\frac{(\alpha + \beta)^2}{2} + \sum_{j=2}^m \left(d_j \cdot \left(\alpha + \beta + \sum_{k=2}^{j-1} d_k + \frac{d_j}{2} \right) \right) = \frac{1}{8} .$$

Thus,

$$\sum_{j=2}^m \left(d_j \cdot \left(\alpha + \beta + \sum_{k=2}^{j-1} d_k + \frac{d_j}{2} \right) \right) = \frac{1}{8} - \frac{(\alpha + \beta)^2}{2} . \quad (14)$$

Similarly, if $y \in [\hat{x}_i, \hat{x}_i^*]$, we apply Lemma B.4 with $y_1 = x_i, y_2 = \hat{x}_i^*, y_{m+1} = \hat{x}_i, y_j = \hat{x}_{n-j+3}$ for $j = 3, \dots, m+1$, and $d_1 = \alpha + \beta$. We get the same expected cost as in Equation (14).

Taken together,

$$\text{cost}(\text{rc}(\mathbf{x}^* \cup \{x_i\}), x_i) = 2 \left(\frac{1}{8} - \frac{(\alpha + \beta)^2}{2} \right) = \frac{1}{4} - (\alpha + \beta)^2 .$$

It holds that

$$\frac{1}{4} - (\alpha + \beta)^2 \geq \frac{1}{4} - \frac{\alpha}{2} - \frac{\beta}{2} \Leftrightarrow \alpha + \beta \leq \frac{1}{2},$$

but the last inequality follows directly from the fact that $x_i \in [r, l]$. This establishes Equation (13), and thus completes the proof of Lemma B.8. \square

Lemma B.9 (Not semicircle to not semicircle). *Assume that $\mathbf{x} \in G^n$ is such that the agents are not on one semicircle, and agent i deviates such that in the new location profile \mathbf{x}' the agents are not on one semicircle. Then*

$$\text{cost}(\text{rc}(\mathbf{x}), x_i) \leq \text{cost}(\text{rc}(\mathbf{x}'), x_i) .$$

Proof. The proof of the lemma follows quite directly from the previous lemmata. Indeed, by Lemma B.7 we have that

$$\text{cost}(\text{rc}(\mathbf{x}' \cup \{x_i\}), x_i) \leq \text{cost}(\text{rc}(\mathbf{x}'), x_i) .$$

Let $p, q \in G$ be the two antipodal points adjacent to x_i . By Lemma B.4, for any $\hat{x}'_i \in G \setminus [p, q]$,

$$\text{cost}(\text{rc}(\mathbf{x}), x_i) = \text{cost}(\text{rc}(\mathbf{x}' \cup \{x_i\}), x_i) .$$

Hence, it sufficient to handle the case where $\hat{x}'_i \in [p, q]$. Notice that $x_i \in [p, q]$, that is, x_i is in the *short* arc between p and q , by our assumption that the agents in \mathbf{x} are not on one semicircle. Therefore, we can apply Lemma B.3 to this arc (by normalizing its length, and replacing x in the lemma by x_i and y by \hat{x}'_i). It follows that the optimal location for \hat{x}'_i is the edges of the arc $[p, q]$, which directly means that

$$\text{cost}(\text{rc}(\mathbf{x}), x_i) \leq \text{cost}(\text{rc}(\mathbf{x}' \cup \{x_i\}), x_i) .$$

\square

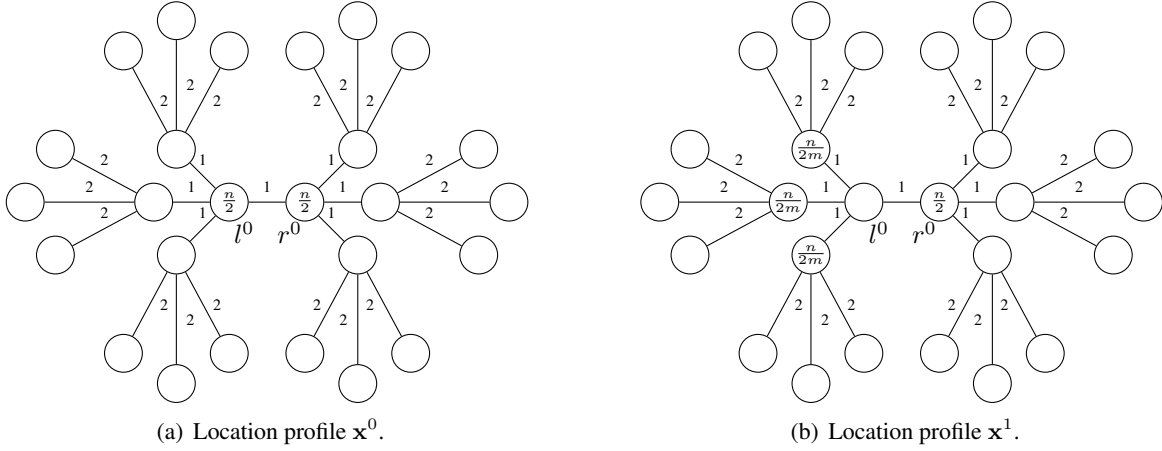


Figure 7: An Illustration of the proof of Theorem 4.2, for $m = 3$ and $k = 2$. A number inside a node indicates the number of agents that are located at this node.

C Proof of Theorem 4.2

Let $m, k \in \mathbb{N}$, whose value remains to be determined. The construction of the graph G is recursive, and depends on m and k . We start with an edge of length 1, which connects the vertices l^0 (for “left”) and r^0 (for “right”). The vertex l^0 is connected to m vertices via edges of length 1; these vertices are called *left vertices on level 1*. Each left vertex on level 1 is connected to m vertices via m edges of length 2; these vertices are called *left vertices on level 2*. In general, each left vertex on level d is connected to m left vertices on level $d + 1$ via edges of length 2^d . The maximum level is k , that is, the left vertices on level k are leaves. The construction is symmetric with respect to the right vertices, i.e., r^0 is connected to m right vertices on level 1 via edges of length 1, and so on.

Now, let $f : G^n \rightarrow \Delta(G)$ be a randomized SP mechanism. Assume for ease of exposition that $2m^k$ divides n , and consider a location profile \mathbf{x}^0 where there are $n/2$ agents at l^0 and $n/2$ agents at r^0 ; this profile is illustrated in Figure 7(a). Clearly, since the distance between l^0 and r^0 is 1, we have that $\mathbb{E}[d(f(\mathbf{x}^0), l^0)] \geq 1/2$, or $\mathbb{E}[d(f(\mathbf{x}^0), r^0)] \geq 1/2$; assume without loss of generality that the former statement is true.

Next, consider the location profile \mathbf{x}^1 , where we have $n/2m$ agents in each left vertex on level 1, and $n/2$ agents in r^0 ; this profile is illustrated in Figure 7(b). We claim that it still holds that $\mathbb{E}[d(f(\mathbf{x}^0), l^0)] \geq 1/2$. Indeed, \mathbf{x}^1 can be obtained from \mathbf{x}^0 by moving the agents, one by one, from l^0 to the left vertices on level 1. Due to strategyproofness, the expected distance from l^0 cannot decrease after each deviation. Since the expected distance was initially at least $1/2$, this is true after the $n/2$ agents have all deviated.

Denote the left vertices on level 1 by l_1^1, \dots, l_m^1 . For any point $y \in G$ such that $d(y, l^0) \geq 1/2$, we have that

$$\sum_{j=1}^m d(y, l_j^1) \geq m + (m - 2) \cdot \frac{1}{2} = \frac{3m - 2}{2} .$$

Since this is true point-wise, it is also true that the sum of expected distances between $f(\mathbf{x}^1)$ and the left vertices on level 1 is at least $\frac{3m-2}{2}$, that is,

$$\sum_{j=1}^m \mathbb{E}[d(f(\mathbf{x}^1), l_j^1)] \geq \frac{3m - 2}{2} .$$

By averaging over these m vertices we conclude that there exists a left vertex on level 1, denoted l^1 , such that $\mathbb{E}[d(f(\mathbf{x}^1), l^1)] \geq 3/2 - 1/m$.

We subsequently consider the location profile \mathbf{x}^2 that is obtained from \mathbf{x}^1 by moving the $n/(2m)$ agents from l^1 to its m neighbors on level 2, such that each left vertex on level 2 that is adjacent to l^1 has $n/(2m^2)$ agents. By similar arguments as before, we get that $\mathbb{E}[d(f(\mathbf{x}^2), l^1)] \geq 3/2 - 1/m$, and, therefore, that there exists a left vertex on level 2 that is adjacent to l^1 , call it l^2 , such that

$$\mathbb{E}[d(f(\mathbf{x}^2), l^2)] \geq \frac{2m + (m-2)\left(\frac{3}{2} - \frac{1}{m}\right)}{m} \geq \frac{7}{2} - \frac{4}{m}.$$

We inductively build location profiles $\mathbf{x}^3, \mathbf{x}^4, \dots, \mathbf{x}^k$ in this fashion. We then have the following claim.

Lemma C.1. *There exists a left vertex on level k , denoted l^k , such that*

$$\mathbb{E}[d(f(\mathbf{x}^k), l^k)] \geq \frac{2^{k+1} - 1}{2} - \frac{2^{k+1} - (k+2)}{m}. \quad (15)$$

Proof. We prove the lemma by induction on the level d . For $d = 0$, we have that

$$\frac{2^{0+1} - 1}{2} - \frac{2^{0+1} - (0+2)}{m} = \frac{1}{2},$$

which is indeed the lower bound that we have obtained for $\mathbb{E}[d(f(\mathbf{x}^0), l^0)]$.

Let us assume that there exists a left vertex on level d , denoted l^d , such that

$$\mathbb{E}[d(f(\mathbf{x}^d), l^d)] \geq \frac{2^{d+1} - 1}{2} - \frac{2^{d+1} - (d+2)}{m}.$$

Constructing \mathbf{x}^{d+1} along the lines given above, we get that there is a left vertex on level $d+1$, denoted l^{d+1} , such that

$$\begin{aligned} \mathbb{E}[d(f(\mathbf{x}^{d+1}), l^{d+1})] &\geq \frac{1}{m} \left[2^d \cdot m + (m-2) \left(\frac{2^{d+1} - 1}{2} - \frac{2^{d+1} - (d+2)}{m} \right) \right] \\ &\geq \frac{2^{d+2} - 1}{2} - \frac{2^{d+2} - (d+3)}{m}. \quad \square \end{aligned}$$

From Lemma C.1 we obtain a location profile \mathbf{x}^k and a left vertex on level k , l^k , such that Equation (15) holds, hence the expected maximum distance is at least the expression at the right hand side of Equation (15). On the other hand, under \mathbf{x}^k the solution that locates the facility at l^{k-1} has a maximum cost of 2^{k-1} . The ratio is at least

$$\frac{\text{mc}(f(\mathbf{x}^k), \mathbf{x}^k)}{\text{mc}(l^{k-1}, \mathbf{x}^k)} \geq \frac{\frac{2^{k+1}-1}{2} - \frac{2^{k+1}-(k+2)}{m}}{2^{k-1}} \geq 2 - \frac{1}{2^k} - \frac{4}{m}.$$

Note that we have to choose m and k such that $n/(2m^k) \geq 1$, so that we still have at least one agent at each of m left vertices on level k in our construction. By taking $k = \Theta(\sqrt{\log n})$, and

$$m = \Theta\left(n^{1/k}\right) = \Theta\left(n^{\frac{1}{\sqrt{\log n}}}\right),$$

we satisfy the above constraint and get that the approximation ratio of f is as announced. \square