Ordinal Approximation for Social Choice, Matching, and Facility Location Problems Given Candidate Positions

ELLIOT ANSHELEVICH and WENNAN ZHU, Rensselaer Polytechnic Institute

9

In this work, we consider general facility location and social choice problems, in which sets of agents $\mathcal{A}$ and facilities $\mathcal{F}$ are located in a metric space, and our goal is to assign agents to facilities (as well as choose which facilities to open) to optimize the social cost. We form new algorithms to do this in the presence of only *ordinal information*, i.e., when the true costs or distances of the agents from the facilities are *unknown*, and only the ordinal preferences of the agents for the facilities are available. The main difference between our work and previous work in this area is that, while we assume that only ordinal information about agent preferences is known, we also know the exact locations of the possible facilities $\mathcal{F}$. Due to this extra information about the facilities, we are able to form powerful algorithms that have small *distortion*, i.e., perform almost as well as omniscient algorithms (which know the true numerical distances between agents and facilities) but use only ordinal information about agent preferences. For example, we present natural social choice mechanisms for choosing a single facility to open with distortion of at most 3 for minimizing both the total and the median social cost; this factor is provably the best possible. We analyze many general problems including matching, k -center, and k -median, and we present black-box reductions from omniscient approximation algorithms with approximation factor β to ordinal algorithms with approximation factor $1 + 2\beta$; doing this gives new ordinal algorithms for many important problems and establishes a toolkit for analyzing such problems in the future.

CCS Concepts: • **Theory of computation** → **Design and analysis of algorithms**;

Additional Key Words and Phrases: Social choice, matching, distortion, ordinal information, approximation algorithms, facility location

ACM Reference format:

Elliot Anshelevich and Wennan Zhu. 2020. Ordinal Approximation for Social Choice, Matching, and Facility Location Problems Given Candidate Positions. *ACM Trans. Econ. Comput.* 9, 2, Article 9 (January 2021), 24 pages.

<https://doi.org/10.1145/3434417>

1 INTRODUCTION

Many important problems involve assigning agents to facilities. For example, assigning patients to hospitals, students to universities, people to houses, and so on. The target of assignment problems is usually to minimize social cost or maximize social welfare. When we consider the social cost of assignment problems, it is natural to assume that the agents prefer facilities that are “closer” to

This work was partially supported by NSF Award No. CCF-1527497.

Authors' addresses: E. Anshelevich and W. Zhu, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, 12180; emails: eanshel@cs.rpi.edu, zhuw5@rpi.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

2167-8375/2020/01-ART9 \$15.00

<https://doi.org/10.1145/3434417>

them in some sense, thus the cost of an agent is often represented by the distance between the agent and the facility it is assigned to. Besides the cost of distances, there are many other cost functions and constraints for different problems; for example, in the capacitated facility assignment problem, each facility has a maximum number of agents it can accommodate.

In this work, we consider general facility location problems, in which sets of agents $\mathcal{A}$ and facilities $\mathcal{F}$ are located in a metric space, and our goal is to assign agents to facilities (as well as choose which facilities to open) so that agents are assigned to facilities that are close to them. For example, $\mathcal{F}$ may consist of the possible locations to open new stores, and the goal may be that all agents have a store near them, or that the sum of agent distances to the stores they are assigned to is small, and so on. This setting also captures many social choice problems, in which the facilities correspond to candidates, and the goal would be to choose a single candidate (and assign all agents to this candidate) so that the distances of the agents from the chosen candidate are small. Here the distances correspond to *spatial preferences*, i.e., the metric space represents the ideological space in which a more preferred candidate would be closer to an agent; see References [4, 24] for a discussion of such spatial preferences in social choice. Our setting also captures matching and many related problems, in which we would open all facilities, but are only able to assign one agent to each facility, thus forming a matching between agents and facilities; facilities here could correspond to houses or items, for example.

If the distances between agents and facilities are known, then we can calculate the optimal solution for these assignment problems. Note that many of the facility location problems are NP-Complete, but at least it is possible to compute optimal assignments of agents to facilities (or the optimal candidates to select for social choice settings) given unlimited computational resources. For many of the settings we mentioned above, however, it is unlikely that we know the exact distances of the agents from the facilities. For social choice these distances would correspond to the cardinal preferences of voters for candidates, for example, “My cost for candidate X winning is exactly 2.35.” It is far more common that only *ordinal* preferences of the agents for the candidates are known, i.e., “I prefer X to Y.” Similarly, when trying to form a matching, or even in general facility location problems where we survey the agents to find out their preferences, it is much easier to elicit ordinal preferences (“I prefer to be matched with X over Y”) over precise numerical preferences. These observations have recently led to a large body of work using the *utilitarian approach*, in which we assume that some latent numerical costs or utilities exist, but we only know the *ordinal* preferences of the agents, not their underlying numerical costs. See, for example, References [4, 6, 16, 21, 28, 33, 45] for the social choice setting, References [1, 7–9] for matching and other graph problems, and Reference [18] for facility location. These works focus on measuring the *distortion* of various algorithms: a measure of how well an algorithm behaves when using only ordinal information, as compared to the optimal algorithm that has access to the true underlying numerical information. More formally, the *distortion* [4, 44] of an assignment is defined as the worst-case ratio of its social cost to the social cost of the optimal solution.

As in the work mentioned above, we assume that only ordinal information about the distances between agents and facilities is known. However, although the locations and numerical preferences of the agents are usually difficult to obtain, the locations of facilities are mostly public information. The locations of political candidates in ideological space can be reasonably well estimated based on their voting records and public statements. When forming a survey about new stores to open, we may not know exactly how much the customers would prefer one store over the other, since the customer locations may be private, but the locations of the possible stores themselves are public knowledge. The main difference between our work and previous work in this area is that we assume:

While only ordinal information about agent preferences is known, we know the exact locations of the possible facilities $\mathcal{F}$.

As we discuss below, this extra information about the locations of the facilities relative to each other allows us to produce much stronger algorithms, and show much nicer bounds on distortion. Note that this information might be available in the setting of other related papers, but not explicitly exploited. In fact, in many cases, we do not even need the full information about the locations of the facilities. The main message of this article is that having a small amount of information about the candidates in social choice settings, or the facilities in facility location, allows us to obtain solutions that are provably *close to optimal* for a large class of problems even though the only information we have about the agent preferences is ordinal, and thus it is impossible (even given unlimited computational resources) to compute the *true* optimal solution.

Our Contributions. We begin by looking at the social choice setting, in which we have a set of n agents $\mathcal{A}$ and m candidates $\mathcal{F}$ in a metric space, and we are given an ordinal ranking of each agent for the candidates. This setting was considered in, e.g., References [4, 6, 21, 28, 33, 35, 45]. In particular, for the objective of minimizing the total distance of the agents from the chosen candidate, Reference [4] showed that Copeland and similar voting mechanisms always have distortion of at most 5. The best known deterministic algorithm is given by a recent paper [42] with a distortion of 4.236. While Reference [4] showed that no deterministic voting mechanism can achieve a worst-case distortion of less than 3, it is still an open question to close the upper and lower bounds. In this article, we show that if we know the exact locations of the candidates in addition to the ordinal ranking of the agents, then there is a simple algorithm that achieves a distortion of 3, and no better bound is possible. In other words, while we do not know the true distances from agents to candidates, we can compute an outcome that is a 3-approximation *no matter what* the true distances are, as long as they are consistent with the ordinal preferences given to us. Moreover, this approximation is possible even if for each agent we are only given their favorite (i.e., top-choice) candidate: there is no need for the agents to submit a full preference ranking over all alternatives.

We also study other objective functions in addition to minimizing the total distance from agents to the chosen alternative. We give a natural deterministic voting mechanism that has distortion at most 3 for objectives such as minimizing the median voter cost, the egalitarian objective of minimizing the maximum voter cost, and many other objectives. This mechanism achieves all these approximation guarantees *simultaneously*, and moreover it does not need the exact locations of the candidates: it suffices to be given an ordinal ranking of the distances from each candidate to each other candidate. In other words, this mechanism is especially suitable for the case when candidates are a subset of voters, as our mechanism will obtain the ordinal ranking of each voter for all candidates, and this is the only information that would be required. Note that Reference [4] proved that *no* deterministic mechanism can achieve a distortion better than 5 for the median objective; the reason why we are able to achieve a distortion of 3 here is precisely because we also know how each candidate ranks all the other candidates, in addition to how each voter ranks all candidates.

We then proceed to our general facility assignment model. We are given a set of agents and a set of facilities in a metric space. The distances between facilities are given, but the distances between agents and facilities are unknown; instead, we only know ordinal preferences of the agents over the facilities that are consistent with the true underlying distances. There could be arbitrary constraints on the assignment, such as facility capacities, or constraints enforcing that some agents cannot be (or must be) assigned to the same facility, and so on. A valid assignment is to assign each agent to a facility without violating the constraints. We consider many different social cost

Table 1. Best Known Distortion of Polynomial-time Algorithms in Different Settings

	Omniscient: full distances	Agents' ordinal prefs and facility locations	Only agents' ordinal prefs (lower bounds)
Total (Sum) Social Choice	1	3	4.236 (3)
Median Social Choice	1	3	5 (5)
Min Weight Bipartite Matching	1	3	n (3)
Egalitarian Bipartite Matching	1	3	- (2)
Facility Location	1.488 [40]	3.976	∞ (∞)
k -center	2 [36]	5	- (-)
k -median	2.675 [17]	6.35	- ($\Omega(n)$)

“Omniscient” stands for the setting where all distances between agents and facilities are known, and the numbers represent the best-known approximation ratios. The second column represent our setting, in which the ordinal preferences of the agents, and the numerical distances between facilities are known. The last column represents the pure ordinal setting in which only the agent ordinal preferences are known, but the distances between facilities are unknown; this setting has been previously studied, and we include the known lower bounds on the possible distortion in parentheses, including some that we prove in the Appendix.

functions to optimize. For a general class of cost functions (essentially ones that are monotone and subadditive), we give a black-box reduction that converts an algorithm for the omniscient version of this problem (i.e., the version where the true distances are known) to an ordinal algorithm with small distortion. Specifically, if we have an omniscient algorithm that always produces an assignment that is a β -approximation of the optimum, then using it, we can create an ordinal algorithm that only knows the ordinal preferences of the agents instead of their true distances to the facilities, but has distortion of at most $1 + 2\beta$.

Many well-known problems fall into our facility assignment model; Table 1 summarizes some of our results. For example, classic facility location with facility costs, minimum weight bipartite matching, egalitarian bipartite matching, k -center, and k -median are all special cases. In particular, our results show that if we are given unbounded computational resources, then it is always possible to form an assignment with distortion of at most 3 for these problems, and no better bound is possible simply due to the fact that we do not possess all the relevant information to compute the true optimum. This is a large improvement over previously known distortion bounds: for minimum cost ordinal matching the best-known distortion bound is n using random serial dictatorship [18]; by using the knowledge of facility locations, we are able to reduce this approximation ratio to 3.

Discussion and Related Work. Ordinal approximation for the minimum social cost (or maximum social welfare) with underlying utilities/distances between agents and alternatives has been studied in many settings including social choice [4, 6, 11, 13, 15, 16, 19, 21, 25, 28, 30, 31, 33, 38, 39, 41, 43–45], matchings [7–9, 12, 18, 22, 29], secretary problems [37], participatory budgeting [10, 34], general graph problems [1, 7], hierarchical clustering [23], and many other models in recent years. The general assumption of the ordinal setting is that we only have the ordinal preferences of the agents over the alternatives, and the goal is to form a solution that has close to optimal social cost. There are different models: social choice, matching, facility location, and so on; different objectives: minimizing social cost, maximizing social welfare, total cost objective, median objective, egalitarian objective, and so on; different assumptions on utility or cost functions: unit-sum, unit-range, metric space, and so on. In this article, we study general facility assignment problems in a metric space, and assume that the ordinal preferences of the agents over the alternatives are given. Unlike previous work on this topic, we also assume the locations of the alternatives are known; we show that this extra information enables us to achieve much better approximation ratios than in the pure ordinal setting for many problems.

The distortion of social choice functions was first introduced in Reference [44] to describe the ratio between the total utility of the optimal candidate and the candidate selected by a mechanism using only the ordinal preferences. References [4, 33, 45] studied the distortion of social choice functions in a metric space; the assumption that the underlying numerical costs have this metric property allows for much better results than more general costs. In particular, for the objective of minimizing the total of the agents from the chosen candidate, the above papers were able to show good distortion bounds for many well-known mechanisms, in particular a bound of 5 for Copeland [4], a bound of $O(\ln m)$ for Single Transferable Vote (STV) [45], and many others. In addition, Reference [4] proved that no deterministic mechanism can have worst-case distortion better than 3, and Reference [45] showed that all scoring rules for m -candidates have a distortion of at least $1 + 2\sqrt{\ln m - 1}$. Abramowitz et al. [2] studied distortion with voters' preference strength in a metric space and Amanatidis et al. [3] also studied preference strength with unit-sum or unrestricted utility functions. Goel et al. [33] showed that Ranked Pairs, and the Schulze rule have a worst-case distortion of at least 5, and the expected worst-case distortion of any (weighted)-tournament rule is at least 3. They also introduced the notion of "fairness" of social choice rules, discussed the fairness ratio of Copeland, Randomized Dictatorship, a general class of cost functions, and studied the relationship between fairness and distortion [32]. Recently, Munagala et al. [42] proposed a deterministic algorithm that improved the distortion to 4.236, which is the best known approach. In this article, we show that if we know the exact locations of the candidates in addition to the ordinal ranking of the agents, then there is a simple algorithm that achieves a distortion of 3, and no better bound is possible.

While the above work, as well as our article, only focuses on deterministic algorithms, the distortion of randomized algorithms in social choice has also been considered, see, for example, References [6, 26, 28, 35]. In a slightly different flavor of result, References [20, 21] consider the special case where candidates are randomly and independently drawn from the set of voters. While we leave the analysis of randomized algorithms that know the location of the facilities to future work, and consider the worst-case candidate locations, it is worth pointing out that our *deterministic* algorithm achieves a distortion of 3, which is also the best known distortion bound for any *randomized* mechanism that only knows the ordinal preferences of the agents. Similarly, another common goal is to form *truthful* mechanisms with small distortion for matching and social choice, as in References [8, 18, 28]; we focus on general mechanisms in this article to understand the limitations of knowing only certain kinds of ordinal information, and leave the goal of forming truthful mechanisms for future work.

For the median objective of social choice problems, Reference [4] showed that Copeland gives a distortion of at most 5, while *no* deterministic mechanism can achieve a distortion better than 5. Reference [6] also gave a randomized algorithm that has a distortion of at most 4. In this article, we are able to improve this bound to a tight worst-case distortion of 3 by a deterministic mechanism, because we also know how each candidate ranks all the other candidates, in addition to how each voter ranks all candidates.

The distortion of matching in a metric space has received far less attention than social choice questions. References [7–9] analyzed maximum-weight metric matching; the maximization objective makes this problem far easier, and even choosing a uniformly random matching yields a distortion of a small constant. This is very different from our goal of computing a minimum-cost matching, for which no ordinal approximations better than $\Omega(n)$ are known. Reference [18] studied facility assignment problems in a metric space; they considered the problem with or without resource augmentation, and the cases without augmentation are exactly the minimum weight bipartite matching problem. Reference [18] showed that the approximation ratio of random serial

dictatorship (RSD) is at most n , and gave a lower bound of $2^n - 1$ for the approximation ratio of serial dictatorship (SD), and a lower bound of $n^{0.26}$ for RSD. Their results are the best known ordinal approximations for this problem. In this article, we are able to give a tight 3-approximation for the minimum weight matching problem, given the locations of facilities in addition to the agents' ordinal preferences.

2 DISTORTION OF SOCIAL CHOICE MECHANISMS

For the social choice problems studied in this article, we let $\mathcal{A} = \{1, 2, \dots, n\}$ be a set of agents, and let $\mathcal{F} = \{F_1, F_2, \dots, F_m\}$ be a set of alternatives, which we will also refer to sometimes as *candidates* or *facilities*. We will typically use i and j to refer to agents and W, X, Y, Z to refer to alternatives. Let $\mathcal{S}$ be the set of total orders on the set of alternatives $\mathcal{F}$. Every agent $i \in \mathcal{A}$ has a preference ranking $\sigma \in \mathcal{S}$; by $X \succ_i Y$, we will mean that X is preferred over Y in ranking σ . Although we assume that each agent has a total order of preference over the alternatives and that this order is known to us, for many of our results it is only necessary that the top choice of each agent is known. We say X is i 's top choice if i prefers X to every other alternative in $\mathcal{F}$. We call the vector $\sigma = (\sigma_1, \dots, \sigma_n) \in \mathcal{S}^n$ a preference profile. We say that an alternative X pairwise defeats Y if $|\{i \in \mathcal{A} : X \succ_i Y\}| > \frac{n}{2}$. The goal is to choose a single winning alternative.

Cardinal Metric Costs. In this work, we take the utilitarian view and assume that the ordinal preferences σ are derived from underlying (latent) cardinal agent costs. Formally, we assume that there exists an arbitrary metric $d : (\mathcal{A} \cup \mathcal{F})^2 \rightarrow \mathbb{R}_{\geq 0}$ on the set of agents and alternatives. The cost incurred to agent i by alternative X being selected is represented by $d(i, X)$, which is the distance between i and X . Such spatial preferences are relatively common and well-motivated, see, for example, References [4, 24] and the references therein. The underlying distances $d(i, X)$ are *unknown*, but unlike most previous work, we *do* assume the distances between *alternatives* are given. The distance between two alternatives X and Y is denoted by $\ell(X, Y)$.

The metric costs d naturally give rise to a preference profile. We say that d is *consistent* with σ if $\forall i \in \mathcal{A}, \forall X, Y \in \mathcal{F}$, if $d(i, X) < d(i, Y)$, then $X \succ_i Y$. It means that the cost of X is less than the cost of Y for agent i , so agent i prefers X over Y . As described above, we know exactly the distances ℓ and the preferences σ , but do not know the true costs d that give rise to σ . Let $\mathcal{D}(\sigma, \ell)$ be the set of metrics that are consistent with σ and ℓ ; we know that one of the metrics from this possibly infinite space captures the true costs, but do not know which one. Note that it is proved in Reference [5] that for any preference profile σ , there exists a metric consistent with σ , and this conclusion is still true with candidates' ordinal preference profiles for each other, because all the candidates are points on a simplex in the metric construction in Reference [5].

Social Cost Distortion. We study several objective functions for social cost in this article. First, the most common notion of social cost is the sum objective function, defined as $SC_{\Sigma}(X, \mathcal{A}) = \sum_{i \in \mathcal{A}} d(i, X)$. We also study the median objective function, $SC_{\text{med}}(X, \mathcal{A}) = \text{med}_{i \in \mathcal{A}}(d(i, X))$, as well as the egalitarian objective and many others (see Section 2.2). We use the notion of distortion to quantify the quality of an alternative in the worst case, similarly to the notation in References [16, 44]. For any alternative W , we define the distortion of W as the ratio between the social cost of W and the optimal alternative:

$$\begin{aligned} \text{dist}_{\Sigma}(W, \sigma, \ell) &= \sup_{d \in \mathcal{D}(\sigma, \ell)} \frac{SC_{\Sigma}(W, \mathcal{A})}{\min_{X \in \mathcal{F}} SC_{\Sigma}(X, \mathcal{A})}, \\ \text{dist}_{\text{med}}(W, \sigma, \ell) &= \sup_{d \in \mathcal{D}(\sigma, \ell)} \frac{SC_{\text{med}}(W, \mathcal{A})}{\min_{X \in \mathcal{F}} SC_{\text{med}}(X, \mathcal{A})}. \end{aligned}$$

In other words, saying that the distortion of W is at most 3 means that, no matter what the true costs d are (as long as they are consistent with σ and ℓ , which we know), it must be that the social cost of W is within a factor of 3 of the true optimal alternative, which is impossible to compute without knowing the true costs. Because of this, a small distortion value means that there is no need to obtain the true agent costs, and the ordinal information σ (together with information ℓ about the alternatives) is enough to form a good solution.

A social choice function f on $\mathcal{A}$ and $\mathcal{F}$ takes σ and ℓ as input, and returns the winning alternative. We say the distortion of f is the same as the distortion of the winning alternative chosen by f on σ and ℓ . In other words, the distortion of a social choice mechanism f on a profile σ and facility distances ℓ is the worst-case ratio between the social cost of $W = f(\sigma, \ell)$, and the social cost of the true optimal alternative.

2.1 Distortion of Total Social Cost

In this section, we study the sum objective and provide a deterministic algorithm that gives a distortion of at most 3. According to Reference [4], the lower bound on the distortion for deterministic social choice functions with only the ordinal preferences (without knowing ℓ) is 3. This occurs in the simple example with two alternatives that are tied with approximately half preferring each one. No matter which one is chosen, the true optimum could be the other one, and its social cost can be as much as 3 times better. We include this example in the Appendix for completeness. Because the example in Theorem 4 from Reference [4] only has two alternatives, knowing ℓ does not provide any extra information, and thus that example also provides a lower bound of 3 in our setting, although we assume the distances ℓ between facilities are known in this article. Therefore, our mechanism achieves the best possible distortion in this setting. Note that if we only have ordinal preferences of the agents without the distances between facilities, then the best known approach [42] gave a deterministic algorithm that has a distortion of 4.236. Thus, our results establish that by knowing the distances ℓ between alternatives, it is possible to reduce the distortion from 4.236 to 3, and no better deterministic mechanism is possible.

LEMMA 2.1. *Let W, X be alternatives. If $W \succ_i X$, then $d(i, X) \geq \frac{d(X, W)}{2}$ [Lemma 5 in Reference [4]].*

ALGORITHM 1: Algorithm for the minimum total social cost.

Input: Agents $\mathcal{A} = \{1, 2, \dots, n\}$,
 Alternatives $\mathcal{F} = \{F_1, F_2, \dots, F_m\}$,
 Each agent i 's top choice alternative,
 Distances between alternatives, i.e., $\ell(Y, Z), \forall Y, Z \in \mathcal{F}$

Output: The winning alternative W .

Generate a projected set of agents $\tilde{\mathcal{A}}$: for each agent i , create agent $\tilde{i}$ at the location of i 's top choice alternative in the metric space, and denote the set of the new agents as $\tilde{\mathcal{A}} = \{\tilde{1}, \tilde{2}, \dots, \tilde{n}\}$. For each alternative $X \in \mathcal{F}$, calculate the total social cost on $\tilde{\mathcal{A}}$ by choosing X , i.e., $SC_{\Sigma}(X, \tilde{\mathcal{A}}) = \sum_{\tilde{i} \in \tilde{\mathcal{A}}} d(\tilde{i}, X) = \sum_{\tilde{i} \in \tilde{\mathcal{A}}} \ell(\tilde{i}, X)$.

Final Output: Return the alternative W that has the minimum social cost $SC_{\Sigma}(W, \tilde{\mathcal{A}})$.

In Algorithm 1, we generate a set of projected agents as follows: Given agents $\mathcal{A}$, alternatives $\mathcal{F}$, and the preference profile σ , for each agent i denote alternative X_i as i 's top choice. Then, we create a new agent $\tilde{i}$ at the location of X_i in the metric space, as shown in Figure 1(a); consequently, for every $Y \in \mathcal{F}$, $d(\tilde{i}, Y) = d(X_i, Y)$. Denote the set of the new agents as $\tilde{\mathcal{A}} = \{\tilde{1}, \tilde{2}, \dots, \tilde{n}\}$. For any metric d consistent with ℓ , $d(\tilde{i}, Y) = d(X_i, Y) = \ell(X_i, Y)$, so the distances between agents in $\tilde{\mathcal{A}}$ and alternatives in $\mathcal{F}$ are known to us, unlike the true distances between $\mathcal{A}$ and $\mathcal{F}$.

Fig. 1. (a) For each agent, generate a projected agent at the location of its top choice alternative. (b) A figure demonstrating agent i , i 's top choice alternative Y , i 's projected agent $\tilde{i}$ located at Y , the winner W , and the optimal alternative X for the proof of Theorem 2.2.

THEOREM 2.2. *The distortion of Algorithm 1 for minimum total social cost on $\mathcal{A}$ is at most 3.*

PROOF. Let W denote the winning alternative. W has the minimum social cost on the agent set $\tilde{\mathcal{A}}$, so for any alternative Y , it must be that

$$\frac{SC_{\Sigma}(W, \tilde{\mathcal{A}})}{SC_{\Sigma}(Y, \tilde{\mathcal{A}})} = \frac{\sum_{\tilde{i} \in \tilde{\mathcal{A}}} d(\tilde{i}, W)}{\sum_{\tilde{i} \in \tilde{\mathcal{A}}} d(\tilde{i}, Y)} = \frac{\sum_{i \in \mathcal{A}} d(\tilde{i}, W)}{\sum_{i \in \mathcal{A}} d(\tilde{i}, Y)} \leq 1. \quad (1)$$

Let X denote the true optimal alternative for $\mathcal{A}$. We want to get $dist_{\Sigma}(W, \sigma, l)$ by upper bounding the cost incurred by W compared to X :

$$\begin{aligned} \frac{SC_{\Sigma}(W, \mathcal{A})}{SC_{\Sigma}(X, \mathcal{A})} &= \frac{\sum_{i \in \mathcal{A}} d(i, W)}{\sum_{i \in \mathcal{A}} d(i, X)} \\ &\leq \frac{\sum_{i \in \mathcal{A}} (d(i, \tilde{i}) + d(\tilde{i}, W))}{\sum_{i \in \mathcal{A}} d(i, X)} \\ &= \frac{\sum_{i \in \mathcal{A}} d(i, \tilde{i})}{\sum_{i \in \mathcal{A}} d(i, X)} + \frac{\sum_{i \in \mathcal{A}} d(\tilde{i}, W)}{\sum_{i \in \mathcal{A}} d(i, X)}. \end{aligned} \quad (2)$$

The inequality $d(i, W) \leq d(i, \tilde{i}) + d(\tilde{i}, W)$ is due to the triangle inequality, since d is a metric, as shown in Figure 1(b). $\forall i \in \mathcal{A}$, we know that $\tilde{i}$ is located at i 's top choice alternative, so the distance between i and $\tilde{i}$ must be less than (or equal to) the distance between i and any alternative; thus, $d(i, \tilde{i}) \leq d(i, X)$. Summing up for all $i \in \mathcal{A}$, we get that $\frac{\sum_{i \in \mathcal{A}} d(i, \tilde{i})}{\sum_{i \in \mathcal{A}} d(i, X)} \leq 1$. For any agent i such that X is not i 's top choice, suppose alternative Y is i 's top choice, then $\tilde{i}$ has the same location as Y and $d(\tilde{i}, X) = d(X, Y)$. By Lemma 2.1, $d(i, X) \geq \frac{d(X, Y)}{2}$, thus $d(i, X) \geq \frac{d(\tilde{i}, X)}{2}$. For all i that X is i 's top choice, $d(\tilde{i}, X) = 0$, so the inequality $d(i, X) \geq \frac{d(\tilde{i}, X)}{2}$ holds for all $i \in \mathcal{A}$. Together with inequality 1

Fig. 2. Example that with agents' top choices and the distances between the alternatives, the median objective has a distortion of 5.

and 2,

$$\frac{SC_{\Sigma}(W, \mathcal{A})}{SC_{\Sigma}(X, \mathcal{A})} \leq 1 + \frac{\sum_{i \in \mathcal{A}} d(\tilde{i}, W)}{\sum_{i \in \mathcal{A}} \frac{d(\tilde{i}, X)}{2}} = 1 + 2 \frac{\sum_{i \in \mathcal{A}} d(\tilde{i}, W)}{\sum_{i \in \mathcal{A}} d(\tilde{i}, X)} \leq 3. \quad \square$$

2.2 Distortion of Median Social Cost

In this section, we study the median objective function, and provide a deterministic mechanism that gives a distortion of at most 3. Recall that we define the median social cost of an alternative X as $SC_{\text{med}}(X, \mathcal{A}) = \text{med}_{i \in \mathcal{A}}(d(i, X))$. We will refer to this as $\text{med}(X)$ when d and $\mathcal{A}$ are fixed. If n is even, then we define the median to be the $(\frac{n}{2} + 1)$ th smallest value of the distances. Note that no deterministic mechanism that only knows ordinal preferences can have worst-case distortion better than 5 (Theorem 25 in Reference [4]). With known distances between candidates, we are able to provide a natural social choice function with distortion of 3, which is also provably the best possible distortion in our setting (consider the example in Theorem 4 from Reference [4] again). Moreover, our social choice function only uses ordinal information about the alternatives, and not the full distances ℓ ; in particular, as long as we have ordinal preferences of each alternative for each other alternative (and thus a total order of the distances from each alternative to the others), then our mechanism will work properly. Such ordinal information may be easier to obtain than full distances ℓ ; for example, candidates can rank all the other candidates. In particular, given agents with ordinal preferences such that the candidates are a subset of the agents, our mechanism will always form an outcome with small distortion, even if we do not know the distances ℓ .

Note that using only agents' top choices over alternatives and the distances between the alternatives, as Algorithm 1 does for the total social cost objective, is not enough to give a worst-case distortion of 3 for the median objective. Consider the following example in Figure 2: There are four alternatives W, X, Y, Z , the distances between them are: $d(W, Y) = d(Y, X) = d(X, Z) = d(Z, W) = 2$ and $d(W, X) = d(Y, Z) = 4$. Suppose W is the top choice of agents 1 and 2, X is the top choice of agent 3 and 4, Y is the top choice of agent 5 and 6, and Z is the top choice of agent 7 and 8. This graph is symmetric, so we choose an arbitrary alternative as the winner. Suppose we choose W as the winner, and the distances between agents and candidates are: the distances from agents 1, 2 to W are both 100, the distances from agents 1, 2 to X, Y, Z are all 102. The distances from agents 5, 6 to Y, X are all 1, and the distances from agents 5, 6 to W, Z are all 3. The distances from agents 7, 8 to Z, X are all 1, and the distances from agents 7, 8 to Y, W are all 3. The distances from agents 3, 4 to X are both 1, the distances from 3, 4 to Y, Z are both 3, and the distances from 3, 4 to W are both

5. In this example, the median is the distance from fifth-closest agent to the winning alternative. X is the optimal alternative with $\text{med}(W) = 1$, while $\text{med}(W) = 5$ has a distortion of 5.

We will use the following Lemmas from Reference [4] in the proof of our algorithm (the proofs are included in the Appendix for completeness):

LEMMA 2.3. *For any two alternatives W and Y , we have $\text{med}(W) \leq \text{med}(Y) + d(Y, W)$. [Lemma 23 in Reference [4]]*

LEMMA 2.4. *For any two alternatives Y and P , if P pairwise defeats (or pairwise ties) Y , then $\text{med}(Y) \geq \frac{d(Y, P)}{2}$. [Proved in Theorem 24 in Reference [4]]*

LEMMA 2.5. *Let W, Y be alternatives $\in \mathcal{F}$, if W pairwise defeats (or pairwise ties) Y , then $\text{med}(W) \leq 3\text{med}(Y)$. [Proved in Theorem 24 in Reference [4]]*

The main easy insight that we use in the formation of our algorithm comes from the following lemma.

LEMMA 2.6. *For any three alternatives W, Y , and P , if P pairwise defeats (or pairwise ties) Y , and $d(Y, W) \leq d(Y, P)$, then $\text{med}(W) \leq 3\text{med}(Y)$.*

PROOF. By Lemma 2.3, $\text{med}(W) \leq \text{med}(Y) + d(Y, W)$. By Lemma 2.4, $\text{med}(Y) \geq \frac{d(Y, P)}{2}$. And we know that $d(Y, P) \geq d(Y, W)$, thus

$$\begin{aligned} \text{med}(W) &\leq \text{med}(Y) + d(Y, W) \\ &\leq \text{med}(Y) + d(Y, P) \\ &\leq \text{med}(Y) + 2\text{med}(Y) \\ &\leq 3\text{med}(Y). \end{aligned}$$

□

We use a natural Condorcet-consistent algorithm to approximate the minimum median social cost with the agents' preference rankings σ and the ordinal preferences of every alternative over other alternatives. First, create the majority graph $G = (\mathcal{F}, E)$, i.e., a graph with alternatives as vertices and an edge $(Y, Z) \in E$ if Y pairwise defeats or pairwise ties Z . If a Condorcet winner (i.e., an alternative that pairwise defeats all others) exists, then we return it immediately.

Otherwise, we consider each pair of alternatives. By Lemma 2.5, if the edge $(W, Y) \in E$, then $\text{med}(W) \leq 3\text{med}(Y)$. When considering an alternative pair W, Y , if $(W, Y) \notin E$ and we know that there exists another alternative P that meets the conditions of Lemma 2.6, then we add an edge (W, Y) to G . It is not difficult to see that whenever the edge (W, Y) is in our graph, this means that $\text{med}(W) \leq 3\text{med}(Y)$. As we prove below, at the end of this process there always exists at least one alternative that has edges to all the other alternatives, and thus the distortion obtained from selecting it is at most 3, no matter which alternative is the true optimal one.

Note that from the ordinal preferences of alternatives over each other, we can get a partial order of distances between the alternatives. Denote this partial order as $\leq$, i.e., we say that $d(W, Y) \leq d(W, Z)$ if we know that W prefers Y to Z (we do not have information about strict preference). This is the information we have on hand: We only know the partial order of distances between pairs of alternatives that share an alternative in common. Note, however, that if there exists a cycle in this partial order, i.e., $d(Y_1, Y_2) \leq d(Y_2, Y_3) \leq d(Y_3, Y_4) \leq \dots \leq d(Y_k, Y_1) \leq d(Y_1, Y_2)$, then this implies that all the distances in the cycle are actually equal, and thus we can also add the relations $d(Y_1, Y_2) \geq d(Y_2, Y_3) \geq d(Y_3, Y_4) \geq \dots \geq d(Y_k, Y_1) \geq d(Y_1, Y_2)$. Such cycles are easy to detect (e.g., by forming a graph with a node for every alternative pair and then searching for cycles), and thus we can assume that whenever a cycle exists in our partial order, then for every pair of distances $d(W, Y)$ and $d(W, Z)$ in the cycle, we have both $d(W, Y) \leq d(W, Z)$ and $d(W, Y) \geq d(W, Z)$.

ALGORITHM 2: Algorithm for the minimum median social cost.

Input: Agents $\mathcal{A} = \{1, 2, \dots, n\}$,
 Alternatives $\mathcal{F} = \{F_1, F_2, \dots, F_m\}$,
 The majority graph $G = (\mathcal{F}, E)$,
 Ordinal preferences of each alternative over other alternatives,
 Partial order of distances between alternatives.

Output: The winning alternative W .

If there is a Condorcet winner W , **then return** W **as the winner**.

forall alternative pairs W, Y **do**

if $(W, Y) \notin E$ or $(Y, W) \notin E$ **then**

 WLOG, suppose (Y, W) exists, but (W, Y) does not exist.

if there exists an alternative P , such that we have $d(Y, W) \leq d(Y, P)$ in our partial order information, and P pairwise defeats (or ties) Y **then**

 Add edge (W, Y) to E ;

 continue;

end

end

end

There must exist an alternative W such that $\forall Y \in \mathcal{F} - \{W\}, (W, Y) \in E$. Return W as the winner.

LEMMA 2.7. Consider the modified majority graph $G = (\mathcal{F}, E)$ at any point during Algorithm 2. For any edge $(W, Y) \in E$, we have that $\text{med}(W) \leq 3\text{med}(Y)$.

PROOF. By Lemma 2.5, for any edge (W, Y) in the original majority graph, $\text{med}(W) \leq 3\text{med}(Y)$.

Now consider an edge (W, Y) added to E when processing the alternative pair W, Y . It must be the case that there exists an alternative P , such that $d(Y, W) \leq d(Y, P)$ and P pairwise defeats (or ties) Y . By Lemma 2.6, $\text{med}(W) \leq 3\text{med}(Y)$. $\square$

LEMMA 2.8. At the end of Algorithm 2, there must exist an alternative W such that $\forall Y \in \mathcal{F} - \{W\}, (W, Y) \in E$.

PROOF. We prove this lemma by contradiction. Suppose no such alternative W exists. Then for each alternative Y , there is at least one alternative Z , such that only $(Z, Y) \in E$ and $(Y, Z) \notin E$. This is because we start with the majority graph, so at least one edge always exists between every pair. We create another directed graph $G' = (\mathcal{F}, E')$, with E' being all the edges (Z, Y) such that $(Y, Z) \notin E$. Thus, any pair of alternatives in G' have at most one direction of edge between them. And by our assumption, each alternative Y has at least one incoming edge in G' . Since the in-degree of each node is at least 1 in G' , there must be at least one cycle in G' . To see this, one can, for example, take the edge (Y_2, Y_1) coming into Y_1 , then the edge (Y_3, Y_2) coming into Y_2 , and proceed in this way until a cycle is formed. Note that every edge in G' must be in the original majority graph, because if we add an edge when processing a pair of alternatives in our algorithm, that pair must have edges in both directions.

Consider a cycle formed by edges $(Y_1, Y_2), (Y_2, Y_3), \dots, (Y_{k-1}, Y_k), (Y_k, Y_1)$. When processing the alternative pair Y_2, Y_3 in Algorithm 2, we did not add edge (Y_3, Y_2) to E , so it must be the case that no alternative P exists such that $d(Y_2, Y_3) \leq d(Y_2, P)$ and P pairwise defeats (or ties) Y_2 . But we know that Y_1 pairwise defeats (or ties) Y_2 , because edge (Y_1, Y_2) is in the original majority graph. Then the only possibility is we do not know if $d(Y_2, Y_3) \leq d(Y_1, Y_2)$, i.e., either $d(Y_2, Y_3)$ and $d(Y_1, Y_2)$ are incomparable in our partial order, or we only know that $d(Y_2, Y_3) \geq d(Y_1, Y_2)$. They cannot be incomparable, since we have the ordinal preferences of Y_2 for Y_1 and Y_3 , thus our partial order must

state that $d(Y_2, Y_3) \geq d(Y_1, Y_2)$, i.e., Y_2 prefers Y_1 to Y_3 . By the same reasoning, we also get that Y_3 prefers Y_2 to Y_4 , and more generally that Y_i prefers Y_{i-1} to Y_{i+1} for all i , where $Y_0 = Y_k$ and $Y_{k+1} = Y_1$, since it is a cycle. This means that in our partial order, we have that $d(Y_1, Y_2) \leq d(Y_2, Y_3) \leq \dots \leq d(Y_{k-1}, Y_k) \leq d(Y_k, Y_1) \leq d(Y_1, Y_2)$. Recall, however, that this means we know $d(Y_1, Y_2) = d(Y_2, Y_3) = \dots = d(Y_{k-1}, Y_k) = d(Y_k, Y_1)$, and before running Algorithm 2, we detect cycles in the partial order of alternative distances, and add the equality information to the partial order. This means that whenever $d(Y_1, Y_2) \leq d(Y_2, Y_3) \leq \dots \leq d(Y_{k-1}, Y_k) \leq d(Y_k, Y_1) \leq d(Y_1, Y_2)$ exists in our partial order, we also have $d(Y_1, Y_2) \geq d(Y_2, Y_3) \geq \dots \geq d(Y_{k-1}, Y_k) \geq d(Y_k, Y_1) \geq d(Y_1, Y_2)$ in the partial order as well. But this gives us a contradiction, since having $d(Y_2, Y_3) \leq d(Y_1, Y_2)$ in the partial order, combined with the fact that Y_1 pairwise defeats Y_2 , would cause us to add the edge (Y_3, Y_2) in our algorithm, which contradicts the statement that only the edge (Y_2, Y_3) is in the final graph produced by the algorithm, but not (Y_3, Y_2) . Thus, there must exist at least one alternative with edges from it to all the others. $\square$

THEOREM 2.9. *The distortion of Algorithm 2 for minimum median social cost is at most 3.*

PROOF. If there is a Condorcet winner, then by Lemma 2.5, the distortion is at most 3.

Otherwise, by Lemma 2.8, the algorithm always returns a winner. Suppose it returns alternative W as the winner, by Lemma 2.7, W has a distortion at most 3 with any alternative X as the optimal solution. $\square$

2.2.1 Generalizing Median: Percentile Distortion. Instead of just considering the median objective, we also consider a more general objective: the α -percentile social cost. Let α -PC(Y) denote the value from the set $\{d(i, Y) : i \in \mathcal{A}\}$, that α fraction of the values lie below α -PC(Y). Thus, median is a special case when $\alpha = \frac{1}{2}$, $\text{med}(Y) = \frac{1}{2}$ -PC(Y). It was shown in Reference [4] Theorem 28 that the worst-case distortion when $\alpha \in [0, \frac{1}{2})$ in that setting (only have agent's ordinal preferences over alternatives) is unbounded, and the same example shows $\alpha \in [0, \frac{1}{2})$ in our setting is also unbounded. However, we are able to give a distortion of 3 for $\alpha \in [\frac{1}{2}, 1]$ in this article, while for the setting in Reference [4], the lower bound for distortion when $\alpha \in [\frac{1}{2}, \frac{2}{3}]$ is 5. The reason is that the ordinal preferences between alternatives are also available in our setting. We will show that Algorithm 2 gives a distortion of at most 3 not only for the median objective but also for the general α -percentile objective, because all the lemmas we used to prove the conclusion for the median objective could be generalized to α -percentile.

We use the following lemma from Reference [4] in the proof of our algorithm:

LEMMA 2.10. *For any two alternatives W and Y , we have α -PC(W) $\leq$ α -PC(Y) + $d(Y, W)$. [Lemma 29 in Reference [4]]*

We can generalize Lemma 2.6 to the following lemma, and the proof is by using Lemma 2.10 instead of Lemma 2.3 in the proof of Lemma 2.6,

LEMMA 2.11. *For any three alternatives W , Y , and P , if P pairwise defeats (or pairwise ties) Y , and $d(Y, W) \leq d(Y, P)$, then α -PC(W) $\leq$ 3α -PC(Y).*

THEOREM 2.12. *The distortion of Algorithm 2 for the α -PC objective social cost with $\frac{1}{2} \leq \alpha \leq 1$ is at most 3.*

PROOF. Note that Lemma 2.10 is actually a generalization of Lemma 2.3, and Lemma 2.11 is a generalization of Lemma 2.6. Lemma 2.4 generalizes to the α -PC objective, because when $\frac{1}{2} \leq \alpha \leq 1$, for any alternative Y , we know α -PC(Y) $\geq$ $\text{med}(Y)$. Lemma 2.5 generalizes to the α -PC objective (shown in the proof of Lemma 2.5 in the Appendix). Then Lemma 2.7 also generalizes to the α -PC objective, because it only uses Lemma 2.5 and Lemma 2.6 in the proof. And Lemma 2.8 still holds

for the same algorithm. Thus, all the lemmas and properties of the median objective used in the proof of Theorem 2.9 could be generalized into the α -PC objective, so the conclusion still holds for the α -PC objective when $\frac{1}{2} \leq \alpha \leq 1$. $\square$

2.2.2 Algorithm 2 and the Total Social Cost. Although Algorithm 2 is designed for the median objective, it also performs quite well for the sum objective. Interestingly, the distortion of this algorithm for the minimum total social cost is at most 5, which slightly worse than the best known deterministic algorithm (with distortion of 4.236) with no knowledge of candidate preferences. Thus, this algorithm gives a distortion of 3 for median (and in fact for all α -percentile objectives when $\frac{1}{2} \leq \alpha \leq 1$) and distortion of 5 for sum simultaneously. In settings where we are not sure which objectives to optimize, or ones where we care both about the total social good, and about fairness, this social choice mechanism provides the best of both worlds. The lemmas and proofs for this result are similar to Theorem 2.9, as follows.

LEMMA 2.13. *Let W, Y be alternatives $\in \mathcal{F}$. If W pairwise defeats (or pairwise ties) Y , then $SC_{\Sigma}(W, \mathcal{A}) \leq 3SC_{\Sigma}(Y, \mathcal{A})$. [Proved in Corollary 7 in Reference [4]]*

LEMMA 2.14. *For any three alternatives W, Y , and P , if P pairwise defeats (or pairwise ties) Y , and $d(Y, W) \leq d(Y, P)$, then $SC_{\Sigma}(W, \mathcal{A}) \leq 5SC_{\Sigma}(Y, \mathcal{A})$.*

PROOF. For all $i \in \mathcal{A}$, we know $d(i, W) \leq d(i, Y) + d(Y, W)$ by the triangle inequality. Summing up for all $i \in \mathcal{A}$, we get $SC_{\Sigma}(W, \mathcal{A}) \leq SC_{\Sigma}(Y, \mathcal{A}) + n \cdot d(Y, W)$.

P pairwise defeats (or pairwise ties) Y , so at least half of the agents prefer P to Y ; thus the total social cost of Y is at least the sum of the social cost of these half of agents. By Lemma 2.1, we get $SC_{\Sigma}(Y, \mathcal{A}) \geq \frac{n}{2} \frac{d(Y, P)}{2} = \frac{n}{4} d(Y, P)$. Thus,

$$\begin{aligned} SC_{\Sigma}(W, \mathcal{A}) &\leq SC_{\Sigma}(Y, \mathcal{A}) + n \cdot d(Y, W) \\ &\leq SC_{\Sigma}(Y, \mathcal{A}) + n \cdot d(Y, P) \\ &\leq SC_{\Sigma}(Y, \mathcal{A}) + 4SC_{\Sigma}(Y, \mathcal{A}) \\ &\leq 5SC_{\Sigma}(Y, \mathcal{A}). \end{aligned}$$

$\square$

LEMMA 2.15. *Consider the modified majority graph $G = (\mathcal{F}, E)$ at any point during Algorithm 2. For any edge $(W, Y) \in E$, we have that $SC_{\Sigma}(W, \mathcal{A}) \leq 5SC_{\Sigma}(Y, \mathcal{A})$.*

PROOF. By Lemma 2.13, for any edge (W, Y) in the original majority graph, $SC_{\Sigma}(W, \mathcal{A}) \leq 3SC_{\Sigma}(Y, \mathcal{A})$.

Now consider an edge (W, Y) added to E when processing the alternative pair W, Y . It must be the case that there exists an alternative P , such that $d(Y, W) \leq d(Y, P)$ and P pairwise defeats (or ties) Y . By Lemma 2.14, $SC_{\Sigma}(W, \mathcal{A}) \leq 5SC_{\Sigma}(Y, \mathcal{A})$. $\square$

THEOREM 2.16. *The distortion of Algorithm 2 for minimum total social cost is at most 5, and this bound is tight.*

PROOF. If there is a Condorcet winner, then by Lemma 2.13, the distortion is at most 3. Otherwise, suppose the algorithm returns alternative W as the winner; by Lemma 2.15, W has a distortion at most 5 with any alternative X as the optimal solution.

To see that this bound is tight, consider the following example. There are three facilities W, Y , and P . There are q agents who prefer Y to W to P , q agents who prefer P to Y to W , and 1 agent who prefers W to P to Y . We denote these three sets of agents as $\mathcal{A}_Y$, $\mathcal{A}_P$, and $\mathcal{A}_W$ separately. By the preferences of the agents, we know that Y pairwise defeats W , W pairwise defeats P , and P pairwise defeats Y . The distances between facilities are: $d(Y, W) = 2 - 2\epsilon$, $d(W, P) = 2 - \epsilon$, $d(P, Y) = 2$, where ϵ is a very small positive number. $\mathcal{A}_Y$ is located at the same location as Y , so

$d(\mathcal{A}_Y, Y) = 0$, $d(\mathcal{A}_Y, P) = 2$, and $d(\mathcal{A}_Y, W) = 2 - 2\epsilon$. The distances between $\mathcal{A}_P$ and the alternatives are: $d(\mathcal{A}_P, Y) = d(\mathcal{A}_P, P) = 1$, $d(\mathcal{A}_P, W) = 3 - 2\epsilon$. $\mathcal{A}_W$ has a distance of 1 to all alternatives. Run Algorithm 2 on this example, and consider the alternative pair W, Y . Because P pairwise defeats Y and $d(Y, W) \leq d(Y, P)$, we add edge (W, Y) to the graph and make W the winner. The total social cost of W is $q * (2 - 2\epsilon) + q * (3 - 2\epsilon) + 1 = q(5 - 4\epsilon) + 1$. While the optimal solution is to choose Y as the winner, and get a total social cost of $q + 1$. When q is very large and ϵ is very small, the distortion in this example approaches 5. $\square$

3 FACILITY ASSIGNMENT PROBLEMS

The mechanism we used for approximating the total social cost in Theorem 2.2 can be applied to far more general problems. In this section, we describe a set of facility assignment problems that fit in this framework. As before, let $\mathcal{A} = \{1, 2, \dots, n\}$ be a set of agents, and $\mathcal{F} = \{F_1, F_2, \dots, F_m\}$ be a set of facilities, with each agent i having a preference ranking σ_i over the facilities, and $\sigma = (\sigma_1, \dots, \sigma_n)$.

As in the social choice model, we assume that there exists an arbitrary unknown metric $d : (\mathcal{A} \cup \mathcal{F})^2 \rightarrow \mathbb{R}_{\geq 0}$ on the set of agents and facilities. The distances $d(i, F_j)$ between agents and facilities are unknown, but the ordinal preferences σ and the distances ℓ between facilities are given. Let $\mathcal{D}(\sigma, \ell)$ be the set of metrics consistent with σ and ℓ , as defined previously in Section 2.

Unlike for social choice, our goal is now to choose which facilities to open and which agents should be assigned to which facilities. Formally, we must choose an assignment $x : \mathcal{A} \rightarrow \mathcal{F}$, where $x(i)$ is the facility that i is assigned to. Every $i \in \mathcal{A}$ must be assigned to one (and only one) facility in $\mathcal{F}$; other than that, there could be arbitrary constraints on the assignment. Here are some examples of constraints that fall into our framework: each facility F_i has a capacity, which is the maximum number of agents that can be assigned to F_i ; at least (or at most) p facilities should have agents assigned to them; agents i and j must be (or must not be) assigned to the same facility, and so on. The social choice model is a special case of this one with the constraint that exactly one facility must be opened, and all agents must be assigned to it. Note that the constraints are only on the assignment, and independent of the metric space d . An assignment x is valid if it satisfies all constraints. Let $\mathcal{X}$ be the set of all valid assignments.

The cost function of assignments. The cost of an assignment x consists of two parts. The first part is the distance cost between agents and facilities. $\forall i \in \mathcal{A}$, let s_i denote the distance between i and the facility it is assigned to, i.e., $s_i = d(i, x(i))$. For a given metric d and assignment x , let $s(x, d)$ denote the vector of distances between each $i \in \mathcal{A}$ and $x(i)$, i.e., $s(x, d) = (s_1, s_2, \dots, s_n)$. Let $c_d : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}$ be a cost function that takes a vector of distances as input. For example, this could simply sum up all the distances, take the maximum distance for an egalitarian objective, and so on. To be as general as possible, instead of fixing a specific function c_d , we consider the set of distance cost functions that are monotonically nondecreasing and subadditive. Formally, c_d is monotonically nondecreasing means that for any vectors s and s' such that $s \leq s'$ componentwise, we have that $c_d(s) \leq c_d(s')$. Any reasonable cost function should satisfy this property if agents desire to be assigned to closer facilities. c_d being subadditive means that for any vectors s and s' , we have that $c_d(s + s') \leq c_d(s) + c_d(s')$. While not all functions are subadditive, many important ones are, as they represent the concept of “economies of scale,” a common property of realistic costs.

The second part of the assignment cost is the facility cost. Let $c_f(x)$ denote the facility cost for assignment x . c_f can be an *arbitrary* function over the assignments, for example, the opening cost of facilities, the penalty (or reward) for assigning certain agents to the same facility, and so on. Our framework includes all such functions, and thus is quite general, as we discuss below. The

main components needed for our framework to work is that the function c_f does not depend on the distances, only on x , and that the function c_d is subadditive.

The total cost $c(x, d)$ of an assignment x is the sum of the distance cost and the facility cost, i.e. $c(x, d) = c_d(s(x, d)) + c_f(x)$. We study algorithms to approximate the minimum cost assignment given only agents' ordinal preferences over facilities, and the distances between facilities, as described above.

Social Cost Distortion. As for social choice, we use the notion of distortion to measure the quality of an assignment in the worst case, similarly to the notation in References [16, 44]. For any assignment x , we define the distortion of x as the ratio between the social cost of x and the optimal assignment:

$$\text{dist}(x, \sigma, l) = \sup_{d \in \mathcal{D}(\sigma, l)} \frac{c(x, d)}{\min_{x' \in \mathcal{X}} c(x', d)}.$$

A social choice function f on $\mathcal{A}$ and $\mathcal{F}$ takes σ and ℓ as input, and returns a valid assignment on $\mathcal{A}$ and $\mathcal{F}$. We say the distortion of f on σ and ℓ is the same as the distortion of the assignment returned by f . In other words, the distortion of an assignment function f on a profile σ and facility distances ℓ is the worst-case ratio between the social cost of $x = f(\sigma, l)$, and the social cost of the true optimal assignment, which to obtain we would need the true distances d .

Approximation ratio of omniscient algorithms. Consider omniscient algorithms that know the true numerical distances between agents and facilities for the facility assignment problems, in other words, the metric d . In some sense, the goal of our work is to determine when algorithms with only limited information can compete with such omniscient algorithms. With the full distances information, we can of course obtain the optimal assignment using brute force, while for our algorithms with limited knowledge this is impossible even given unlimited computational resources. Nevertheless, we are also interested in what is possible to achieve if we restrict ourselves to polynomial time. To differentiate traditional approximation algorithms from algorithms with small distortion, suppose that an omniscient approximation algorithm $\tilde{f}$ returns assignment x . Then, we denote the approximation ratio of a valid assignment x as

$$\text{ratio}(x) = \frac{c(x, d)}{\min_{x' \in \mathcal{X}} c(x', d)}.$$

Thus, we say the approximation ratio of an omniscient algorithm $\tilde{f}$ is at most β if for any input of the problem, the assignment x returned by $\tilde{f}$ has $\text{ratio}(x) \leq \beta$.

3.1 Examples of Facility Assignment Problems

In this section, we illustrate that our framework is quite general by giving various important examples that fit into our framework. In the section that follows, we prove a general black-box reduction theorem for our framework, and thus immediately obtain mechanisms with small distortion for all these examples simultaneously.

The total social cost problem we discussed in Section 2.1 is a special case of the facility assignment problem such that the constraint is only one facility (alternative) is chosen, and all agents are assigned to it. For any assignment x , the facility cost function $c_f(x) = 0$, and the distance cost function $c_d(s(x, d))$ is the sum of distances from the winning alternative to all agents in the metric d . c_d is monotone and additive (thus subadditive). Here are some other examples that fit in our framework:

Minimum weight metric bipartite matching. Given a set of agents $\mathcal{A}$ and a set of facilities $\mathcal{F}$ such that $|\mathcal{A}| = |\mathcal{F}| = n$. $G = (\mathcal{A}, \mathcal{F}, E)$ is an undirected complete bipartite graph. The facilities

and agents lie in a metric space d . The weight of each edge $(i, F) \in E$ is the distance between i and F , $w(i, F) = d(i, F)$. The goal is to find a minimum weight perfect matching of the bipartite graph given only ordinal information. This setting has been studied before, and the best distortion bound known is n [18] given by RSD for the case when only the ordinal preferences σ are known. Our results show that if we also know the distances ℓ between facilities, then even without knowing the distances d between agents and facilities, it is possible to create simple mechanisms with distortion at most 3 (we can show that no better bound is possible for this setting). Thus, having a bit more information about the facilities immediately improves the distortion bound by a very large amount. We show this result by using our facility assignment framework above: the constraint here is that each facility has a capacity of 1, thus a valid assignment is a perfect matching of the bipartite graph. For any assignment x , the facility cost function is $c_f(x) = 0$, and the distance cost function $c_d(s(x, d))$ is the total edge weight in the assignment. c_d is monotone and additive (thus, subadditive).

Egalitarian bipartite matching. With the same bipartite graph as in minimum weight matching problems, the only difference is that the goal of egalitarian bipartite matching is to find a perfect matching such that maximum edge weight (instead of the total weight) in the matching is minimized [14].

The egalitarian bipartite matching problem is the same as minimum weight bipartite matching except the distance cost function $c_d(s(x, d))$ is the maximum edge weight in the assignment. This function is also monotone and subadditive.

Metric Facility Location. In this problem, one is given a set of agents $\mathcal{A}$ and a set of facilities $\mathcal{F}$ such that $|\mathcal{A}| = n$, $|\mathcal{F}| = m$. The facilities and agents lie in a metric space d . Each facility $F_j \in \mathcal{F}$ has an opening cost f_j . Each agent is assigned to a facility; in different versions there may be capacities on the number of agents assigned to a facility, lower bounds on the number of agents assigned to a facility, or various other constraints [27]. The goal is to find a subset of facilities $\hat{\mathcal{F}} \subseteq \mathcal{F}$ to open, so that the sum of opening costs for facilities in $\hat{\mathcal{F}}$ and total distance of the assignment is minimized.

Our framework allows arbitrary constraints on what constitutes a valid assignment, which captures facilities with capacities or lower bounds if needed. For any assignment x , the facility cost function $c_f(x)$ is the sum of the opening costs f_j for those facilities F_j that have at least one agent assigned to it. The distance cost function $c_d(s(x, d))$ is the total distances in the assignment, which is monotone increasing and additive (thus subadditive).

k -center problem. The goal in this classic problem is to open a set of k facilities, with each agent assigned to the closest one. The optimal solution is the subset of $\hat{\mathcal{F}}$ that minimizes $\max_{i \in \mathcal{A}} d(i, x(i))$. To express this in our framework, the constraint is that no more than k facilities have agents assigned to them. For any assignment x , the facility cost function $c_f(x) = 0$, and the distance cost function $c_d(s(x, d))$ is the maximum distance between any agent and facility in the assignment.

k -median problem. This classic problem is the same as k -center, except the goal is to minimize the sum of distances of agents to the facilities instead of the maximum distance.

3.2 Distortion of Facility Assignment Problems

In this section, we study general facility assignment problems, as described in Section 3, and form mechanisms with small distortion. First, we construct a projected problem such that the distances between agents and facilities are known, so it could be solved by an omniscient algorithm. Then, we map the result of the projected problem to the original problem and bound the distortion of the original problem.

Given agents $\mathcal{A} = \{1, 2, \dots, n\}$ and facilities $\mathcal{F} = \{F_1, F_2, \dots, F_m\}$, suppose facility F' is i 's top choice in $\mathcal{F}$. We create a new agent $\tilde{i}$ at the location of F' in the metric space. Consequently, $\forall F \in \mathcal{F}, d(\tilde{i}, F) = d(F', F)$. Denote the set of the new agents as $\tilde{\mathcal{A}} = \{\tilde{1}, \tilde{2}, \dots, \tilde{n}\}$.

The original assignment problem is on agents $\mathcal{A}$ and facilities $\mathcal{F}$, and only the ordinal preferences of the agents in $\mathcal{A}$ over the facilities are given. The projected problem is on agents $\tilde{\mathcal{A}}$ and facilities $\mathcal{F}$, and we know the actual distances between agents in $\tilde{\mathcal{A}}$ and facilities $\mathcal{F}$, since we know the distances ℓ between facilities. The constraints and costs c_d and c_f remain the same for both the original and the projected problem; the only difference is in the distances d . Our main result is that if we have a β -approximation assignment to the minimum assignment cost on the projected problem, then we can get an assignment that has a distortion of $2\beta + 1$ for the original problem in polynomial time.

THEOREM 3.1. *Given a valid assignment $\tilde{x}$ for the projected problem on $\tilde{\mathcal{A}}$ and $\mathcal{F}$, with $\text{ratio}(\tilde{x}) \leq \beta$, the assignment $x(i) = \tilde{x}(\tilde{i})$ has distortion at most $(1 + 2\beta)$ for the original assignment problem on $\mathcal{A}$ and $\mathcal{F}$.*

PROOF. First, $\tilde{x}$ is a valid assignment for the projected problem on $\tilde{\mathcal{A}}$ and $\mathcal{F}$, so x must also be a valid assignment for the original problem on $\mathcal{A}$ and $\mathcal{F}$. This is because the constraints are only on the assignment, and are independent of the metric space d . For the same reason, the facility cost of x equals the facility cost of $\tilde{x}$, $c_f(x) = c_f(\tilde{x})$.

Now consider the distance cost of x . Let x^* denote the optimal assignment for the original problem. $\forall i \in \mathcal{A}$, let $s_i = d(i, x(i))$, $t_i = d(i, \tilde{i})$, $b_i = d(\tilde{i}, x(i))$. Similarly, let $s_i^* = d(i, x^*(i))$, $b_i^* = d(\tilde{i}, x^*(i))$.

For any agent i and facility $x(i)$, by triangle inequality,

$$s_i = d(i, x(i)) \leq d(i, \tilde{i}) + d(\tilde{i}, x(i)) = t_i + b_i.$$

Because c_d is monotonically nondecreasing and subadditive,

$$\begin{aligned} c_d(s_1, s_2, \dots, s_n) &\leq c_d(t_1 + b_1, t_2 + b_2, \dots, t_n + b_n) \\ &\leq c_d(t_1, t_2, \dots, t_n) + c_d(b_1, b_2, \dots, b_n). \end{aligned}$$

Therefore, the cost of our assignment x is bounded as follows:

$$\begin{aligned} c_f(x) + c_d(s(x, d)) &= c_f(x) + c_d(s_1, s_2, \dots, s_n) \\ &= c_f(\tilde{x}) + c_d(s_1, s_2, \dots, s_n) \\ &\leq c_f(\tilde{x}) + c_d(t_1, t_2, \dots, t_n) + c_d(b_1, b_2, \dots, b_n). \end{aligned}$$

Because $\tilde{i}$ is located at i 's top choice facility, and $x^*(i)$ is a facility, we thus know that $t_i \leq s_i^*$, and by monotonicity $c_d(t_1, t_2, \dots, t_n) \leq c_d(s_1^*, s_2^*, \dots, s_n^*)$. Thus,

$$\begin{aligned} c_f(x) + c_d(s(x, d)) &\leq c_f(\tilde{x}) + c_d(t_1, t_2, \dots, t_n) + c_d(b_1, b_2, \dots, b_n) \\ &\leq c_f(\tilde{x}) + c_d(s_1^*, s_2^*, \dots, s_n^*) + c_d(b_1, b_2, \dots, b_n). \end{aligned}$$

We know that $\tilde{x}$ is a β -approximation of the optimal assignment for the projected problem. Its total cost is exactly $c_f(\tilde{x}) + c_d(b_1, b_2, \dots, b_n)$, since the distance from $\tilde{i}$ to $\tilde{x}(\tilde{i}) = x(i)$ is exactly b_i . Now consider another assignment for the projected problem, in which $\tilde{i}$ is assigned to $x^*(i)$. The cost of this assignment is $c_f(x^*) + c_d(b_1^*, b_2^*, \dots, b_n^*)$, by definition of b_i^* . Since $\tilde{x}$ is a β -approximation, we therefore know that

$$c_f(\tilde{x}) + c_d(b_1, b_2, \dots, b_n) \leq \beta c_f(x^*) + \beta c_d(b_1^*, b_2^*, \dots, b_n^*),$$

and thus

$$\begin{aligned} c_f(x) + c_d(s(x, d)) &\leq c_f(\tilde{x}) + c_d(s_1^*, s_2^*, \dots, s_n^*) + c_d(b_1, b_2, \dots, b_n) \\ &\leq c_d(s_1^*, s_2^*, \dots, s_n^*) + \beta c_f(x^*) + \beta c_d(b_1^*, b_2^*, \dots, b_n^*). \end{aligned}$$

For any agent i and facility $x^*(i)$ in x^* , by triangle inequality,

$$b_i^* = d(\tilde{i}, x^*(i)) \leq d(i, x^*(i)) + d(i, \tilde{i}) \leq 2d(i, x^*(i)) = 2s_i^*.$$

$d(i, \tilde{i}) \leq d(i, x^*(i))$ above, since $\tilde{i}$ is located at the closest facility to i . Because c_d is monotone and subadditive, we also have that

$$c_d(b_1^*, b_2^*, \dots, b_n^*) \leq c_d(2s_1^*, 2s_2^*, \dots, 2s_n^*) \leq 2c_d(s_1^*, s_2^*, \dots, s_n^*).$$

Putting everything together,

$$\begin{aligned} c_f(x) + c_d(s(x, d)) &\leq c_d(s_1^*, s_2^*, \dots, s_n^*) + \beta c_f(x^*) + \beta c_d(b_1^*, b_2^*, \dots, b_n^*) \\ &\leq \beta c_f(x^*) + c_d(s_1^*, s_2^*, \dots, s_n^*) + 2\beta c_d(s_1^*, s_2^*, \dots, s_n^*) \\ &= \beta c_f(x^*) + (1 + 2\beta)c_d(s_1^*, s_2^*, \dots, s_n^*) \\ &\leq (1 + 2\beta)(c_f(x^*) + c_d(s(x^*, d))). \end{aligned} \quad \square$$

Note that the above theorem immediately implies that if we are only concerned with what is possible to achieve given limited ordinal information in addition to distances between facilities, and are not worried about our algorithms running in polynomial time, then we can always form an assignment with distortion of at most 3 from knowing only σ and ℓ . This is because we can solve the projected problem with brute force, and then we have $\beta = 1$. This bound of 3 is tight for many facility assignment problems: consider, for example, an instance of min-cost metric matching with two agents and two facilities, with both preferring F_1 to F_2 . One of the agents has distance to F_1 of 0, and one is located halfway between F_1 and F_2 , but since we only have ordinal information, we do not know which agent is which. If we assign the wrong agent to F_1 , then we end up with distortion of 3, and it is impossible to do better for any deterministic mechanism.

If, however, we want to form poly-time algorithms with small distortion, then the above theorem gives a black-box reduction: If we have a β -approximation algorithm for the omniscient case, then we can form a $1 + 2\beta$ -distortion algorithm for the ordinal case. Actually, we get a $1 + 2\beta$ -distortion for the distance cost, and a β -distortion for the facility cost, which is shown in the second-to-last line of the proof for Theorem 3.1. This leads to the following corollaries:

COROLLARY 3.2. *We can achieve the following distortion in polynomial time:*

- (1) *At most 3 for the minimum weight bipartite matching problem.*
- (2) *At most 3 for Egalitarian bipartite matching.*
- (3) *At most 3.976 for the facility location problem (1.488-approximation for the facility cost, and 3.976-approximation for the distance cost).*
- (4) *At most 5 for the k -center problem.*
- (5) *At most 6.35 for the k -median problem.*

PROOF. Min-cost matching and egalitarian matching are poly-time solvable, so $\beta = 1$. For the latter, one can fix the threshold weight t such that every edge chosen should be at most t , and then determine if such a matching exists. Performing a binary search on t gives an efficient algorithm. For facility location, one can use the omniscient algorithm, which is a 1.488-approximation in Reference [40]. For the k -center problem, a greedy algorithm [36] gives a 2-approximation for the setting that agents are a subset of facilities, which is the case in our projected problem. Reference [17] gives a 2.675-approximation omniscient algorithm for the k -median problem

when agents are a subset of facilities, thus it also gives a 2.675-approximation for our projected problem. $\square$

Note that the median function, unlike sum and maximum, is not subadditive, and thus does not fit into our framework. In fact, while both min-cost and egalitarian matching problems have algorithms with small distortion in our setting, the same is not possible for forming a matching where the objective function is the cost of the *median* edge: See the Appendix for a lower bound.

4 CONCLUSION

In this article, we provided two mechanisms to solve different social cost problems. The first one makes use of the distances between facilities and an omniscient algorithm to get a low distortion for general facility assignment problems. The second mechanism is a new voting rule for social choice that simultaneously achieves a distortion of 3 for many objectives, including the cost of the median voter, and a distortion of 5 for the total social cost at the same time. The first mechanism requires the full distances ℓ , but only needs the top choice from each agent. Thus, it puts only a small load on the agents that submit their preferences, but requires the mechanism designer to collect more information about the facilities and their distances to each other. The second mechanism, however, only requires ordinal preference information from the facilities, but needs the full preference ranking from the agents instead of just the top choice. It is especially appropriate for settings in which the candidates or alternatives are agents themselves.

Many open questions remain for our setting. How well can facility location problems be approximated given information about facilities? While we established upper bounds on distortion, we have no lower bounds besides the trivial bound of 3. What about randomized mechanisms, or what if the mechanisms must be truthful? And more generally, exactly what information is enough to guarantee mechanisms with small distortion? While our results show that knowing information about facility locations is enough to result in small distortion, and recent work shows the same for a small number of queries of voters' preference strength [2, 3], it may be possible that obtaining even a bit of other information would result in powerful approximation algorithms. We look forward to future work on this topic.

A BAD EXAMPLES AND LOWER BOUNDS

We include the lower bound example in Theorem 4 from Reference [4] for completeness.

THEOREM A.1. *No deterministic ordinal voting rule has worst case distortion less than 3 for the sum objective of social choice. [Theorem 4 in Reference [4]]*

PROOF. Suppose there are only two candidates X and W , and two agents 1 and 2. Agent 1 prefer X over W and agent 2 prefer W over X . Without loss of generality, suppose the voting rule chooses W as the winner. Consider the following metric: $d(1, X) = 0$, $d(1, W) = 2$, $d(2, X) = 1 + \epsilon$, $d(2, W) = 1 - \epsilon$. Then $SC_{\Sigma}(X) = 1 + \epsilon$ and $SC_{\Sigma}(W) = 3 - \epsilon$. When ϵ approaches 0, the distortion approaches 3. $\square$

Note that our Algorithm 2 is only for social choice problems, and does not fit in the definition of our general facility assignment problems. This is because the median cost function, unlike sum and maximum, is not subadditive. In fact, while both min-cost and egalitarian matching problems have algorithms with small distortion in our setting, the same is not possible for forming a matching where the objective function is the cost of the *median* edge.

THEOREM A.2. *The worst-case distortion of the median-cost bipartite matching problem in a metric space (given both agent preference profiles and distances between facilities) is unbounded.*

PROOF. Consider the following example: there are three agents a, b, c , and three facilities X, Y, Z . The preferences of the agents are: $a, b \in XYZ$, while $c \in ZXY$. The distances between facilities are: $\ell(X, Y) = 2$, $\ell(X, Z) = \ell(Y, Z) = 1,000$. The distances between the agents and facilities are, of course, unknown. Consider the instance $d(c, Z) = \epsilon$, $d(a, X) = 2\epsilon$, and $d(b, X) = d(b, Y) = 1$. ϵ is a very small positive real number, and other distances not given obey triangle inequality. In this instance, the optimal solution is $x^* = \{(a, X), (b, Y), (c, Z)\}$, which gives a median value of 2ϵ . But because a and b have the same preference profile, the instance could also be $d(c, Z) = \epsilon$, $d(b, X) = 2\epsilon$, and $d(a, X) = d(a, Y) = 1$. If we still return the assignment x^* for this instance, then the median would be 1. The distortion is arbitrarily bad when ϵ approaches 0. $\square$

The following Theorems show some of the lower bounds mentioned in Table 1.

THEOREM A.3. *The worst-case distortion for the facility location problem in a metric space (given only agents' preference profiles) is unbounded.*

PROOF. Consider the following example: there are two agents 1, 2, and two facilities X, Y . Agent 1 prefers X to Y , while agent 2 prefers Y to X . The opening costs are: $c_f(X) = 1$, $c_f(Y) = 100$. We can choose to open one facility or both of them.

Case 1. Suppose we only open X . Consider the following distances between the agents and facilities: $d(1, X) = d(2, Y) = 1$, $d(1, Y) = d(2, X) = L$, for some very large ℓ . If we only open X , then the total cost is $> L$. While the optimal solution is to open both X and Y , which has a total cost of 103. The distortion is unbounded.

Case 2. Suppose we only open Y . Consider the same distances as in **Case 1**, then the total cost is also ℓ . And the optimal solution still has a total cost of 103. The distortion is unbounded.

Case 3. Suppose we open both facilities. Consider the following distances between the agents and facilities: $d(1, X) = d(1, Y) = d(2, X) = d(2, Y) = \epsilon$, where ϵ is a very small positive real number. If we open both facilities, then the total cost is $101 + 2\epsilon$. While the optimal solution is to only open X , which has a total cost of $1 + 2\epsilon$. If we increase $c_f(Y)$, then the approximation ratio is unbounded. $\square$

THEOREM A.4. *The worst-case distortion for the k -median problem in a metric space (given only agents' preference profiles) is at least $\Omega(n)$.*

PROOF. Consider the following example: There are three facilities X, Y , and Z . There are q agents who prefer X to Y to Z , q agents who prefer Y to X to Z , and 1 agent who prefers Z to X to Y . We denote these three sets of agents as $\mathcal{A}_X, \mathcal{A}_Y$ and $\mathcal{A}_Z$ separately. Suppose $k = 2$, then we have three choices of the winners:

Case 1. Choose X, Y as the winners. Consider the following distances between agents and facilities: $d(X, Y) = 1$, $d(Y, Z) = d(X, Z) = L$ for some very large ℓ . $\mathcal{A}_X$ is located at the same location as X , $\mathcal{A}_Y$ is at the same location as Y , and $\mathcal{A}_Z$ is at the same location as Z . The cost of choosing X, Y as the winners is ℓ , because we need to assign the agent in $\mathcal{A}_Z$ to X or Y . While the optimal solution is to choose Y, Z as the winners, and get a total cost of q . So the distortion in this case is unbounded.

Case 2. Choose X, Z as the winners. Consider the following distances between agents and facilities: $d(X, Y) = d(Y, Z) = d(X, Z) = 1$, and $\mathcal{A}_X$ locate on top of X , $\mathcal{A}_Y$ locate on top of Y , and $\mathcal{A}_Z$ locate on top of Z . The cost of choosing X, Z as the winner is q , while the optimal solution is to choose X, Y as the winners, and get a total cost of 1. The distortion is q in this case.

Case 3. Choose Y, Z as the winners. Consider the same distances as in **Case 2**. If we choose Y, Z as the winners, then the total cost is still q , and the distortion of this case is also q .

The total number of agents is $n = 2q + 1$, so we can conclude that the distortions in all these three cases are at least $\Omega(n)$. $\square$

THEOREM A.5. *The worst-case distortion of the egalitarian bipartite matching problem in a metric space (given only agents' preference profiles) is at least 2.*

PROOF. Consider the following example: there are two agents 1, 2, and two facilities X, Y . Both agents prefer X to Y . W.L.O.G., assume we match agent 1 to X , and agent 2 to Y . Suppose the distances between agents and facilities are: $d(1, X) = d(1, Y) = 1$, $d(2, X) = \epsilon$, $d(2, Y) = 2$, where ϵ is a very small positive real number. The egalitarian cost of our matching is 2, while the optimal solution is to match agent 1 to Y , and agent 2 to X , which has a cost of 1. $\square$

B PROOFS FROM REFERENCE [4]

In this section, we include the proof of lemmas and theorems from Reference [4] that are used in this article for completeness.

LEMMA 2.1. *Let W, X be alternatives. If $W \succ_i X$, then $d(i, X) \geq \frac{d(X, W)}{2}$. [Lemma 5 in Reference [4]]*

PROOF. Because $W \succ_i X$, we know that $d(i, W) \leq d(i, X)$. By the triangle inequality, $d(X, W) \leq d(i, X) + d(i, W) \leq 2d(i, X)$. $\square$

LEMMA 2.3. *For any two alternatives W and Y , we have $\text{med}(W) \leq \text{med}(Y) + d(Y, W)$. [Lemma 23 in Reference [4]]*

PROOF. By definition, there are at least $\lfloor \frac{n}{2} + 1 \rfloor$ voters with $d(i, Y) \leq \text{med}(Y)$. By the triangle inequality, for each such voter i , $d(i, W) \leq d(i, Y) + d(Y, W)$. Thus, $\text{med}(W) \leq \text{med}(Y) + d(Y, W)$. $\square$

LEMMA 2.4. *For any two alternatives Y and P , if P pairwise defeats (or pairwise ties) Y , then $\text{med}(Y) \geq \frac{d(Y, P)}{2}$. [Proved in Theorem 24 in Reference [4]]*

PROOF. Because P pairwise defeats (or pairwise ties) Y , we know that there are at least $\frac{n}{2}$ voters prefer P to Y . By Lemma 2.1, for each such voter i , $d(i, Y) \geq \frac{d(P, Y)}{2}$. So $\text{med}(Y) \geq \frac{d(Y, P)}{2}$. $\square$

LEMMA 2.5. *Let W, Y be alternatives $\in \mathcal{F}$, if W pairwise defeats (or pairwise ties) Y , then $\text{med}(W) \leq 3\text{med}(Y)$ [Proved in Theorem 24 in Reference [4]]. For the α -PC objective when $\alpha \in [\frac{1}{2}, 1]$, $\alpha\text{-PC}(W) \leq 3\alpha\text{-PC}(Y)$.*

PROOF. We consider the following two cases separately:

Case 1, $\text{med}(W) = \beta d(Y, W)$, $\beta > \frac{3}{2}$. By Lemma 2.3,

$$\frac{\text{med}(W)}{\text{med}(Y)} \leq \frac{\text{med}(W)}{\text{med}(W) - d(W, Y)} = \frac{\beta d(Y, W)}{(\beta - 1)d(Y, W)} = 1 + \frac{1}{\beta - 1}.$$

Because $\beta > \frac{3}{2}$, the result follows in this case.

Case 2, $\text{med}(W) \leq \frac{3}{2}d(Y, W)$. By Lemma 2.4, $\text{med}(Y) \geq \frac{d(Y, W)}{2}$. Thus, the result follows in this case.

To prove this conclusion also holds for the α -PC objective when $\alpha \in [\frac{1}{2}, 1]$, we only need to replace med with $\alpha\text{-PC}$ in the above proof. Note that Lemma 2.4 easily generalizes to $\alpha\text{-PC}$, because for any alternative Y and $\alpha \in [\frac{1}{2}, 1]$, we know that $\alpha\text{-PC}(Y) \geq \text{med}(Y)$. $\square$

LEMMA 2.10. *For any two alternatives W and Y , we have $\alpha\text{-PC}(W) \leq \alpha\text{-PC}(Y) + d(Y, W)$. [Lemma 29 in Reference [4]]*

PROOF. For each agent i with $d(i, Y) \leq \alpha\text{-PC}(Y)$, we know $d(i, W) \leq d(i, Y) + d(Y, W) \leq \alpha\text{-PC}(Y) + d(Y, W)$. Thus, the result follows. $\square$

LEMMA 2.13. Let W, Y be alternatives $\in \mathcal{F}$. If W pairwise defeats (or pairwise ties) Y , then $SC_{\Sigma}(W, \mathcal{A}) \leq 3SC_{\Sigma}(Y, \mathcal{A})$. [Proved in Corollary 7 in Reference [4]]

PROOF. Let $WY = \{i \in \mathcal{A} : W \succ_i Y\}$. W pairwise defeats (or pairwise ties) Y means there are at least $\frac{n}{2}$ voters prefer W to Y , i.e., $|WY| \geq \frac{n}{2}$. Thus,

$$\begin{aligned} \frac{SC_{\Sigma}(W, \mathcal{A})}{SC_{\Sigma}(Y, \mathcal{A})} &= \frac{\sum_{i \in \mathcal{A}} d(i, W)}{\sum_{i \in \mathcal{A}} d(i, Y)} \\ &= \frac{\sum_{i \in WY} d(i, W) + \sum_{i \in YW} d(i, W)}{\sum_{i \in \mathcal{A}} d(i, Y)} \\ &\leq \frac{\sum_{i \in WY} d(i, Y) + \sum_{i \in YW} (d(i, Y) + d(W, Y))}{\sum_{i \in \mathcal{A}} d(i, Y)} \\ &= 1 + \frac{\sum_{i \in YW} d(W, Y)}{\sum_{i \in \mathcal{A}} d(i, Y)} \\ &= 1 + \frac{|YW|d(W, Y)}{\sum_{i \in \mathcal{A}} d(i, Y)}. \end{aligned}$$

By Lemma 2.1, we know $\forall i \in WY, d(i, Y) \geq \frac{d(W, Y)}{2}$, so

$$\sum_{i \in \mathcal{A}} d(i, Y) \geq \sum_{i \in WY} d(i, Y) = |WY| \frac{d(W, Y)}{2}.$$

Remember $|WY| \geq \frac{n}{2}$, combine with the inequalities above,

$$\begin{aligned} \frac{SC_{\Sigma}(W, \mathcal{A})}{SC_{\Sigma}(Y, \mathcal{A})} &\leq 1 + \frac{|YW|d(W, Y)}{\sum_{i \in \mathcal{A}} d(i, Y)} \\ &\leq 1 + \frac{(n - |WY|)d(W, Y)}{|WY| \frac{d(W, Y)}{2}} \\ &= 1 + \frac{2(n - |WY|)}{|WY|} \\ &\leq \frac{2n}{|WY|} - 1 \\ &\leq 3. \end{aligned}$$

□

ACKNOWLEDGMENT

We thank Onkar Bhardwaj for discussion of the lower bound example for the k-median problem.

REFERENCES

- [1] Ben Abramowitz and Elliot Anshelevich. 2018. Utilitarians without utilities: Maximizing social welfare for graph problems using only ordinal preferences. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. 894–901.
- [2] Ben Abramowitz, Elliot Anshelevich, and Wenzan Zhu. 2019. Awareness of voter passion greatly improves the distortion of metric social choice. In *Proceedings of the International Conference on Web and Internet Economics (WINE'19)*.
- [3] Georgios Amanatidis, Georgios Birmpas, Aris Filos-Ratsikas, and Alexandros A. Voudouris. 2020. Peeking behind the ordinal curtain: Improving distortion via cardinal queries. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*.
- [4] Elliot Anshelevich, Onkar Bhardwaj, Edith Elkind, John Postl, and Piotr Skowron. 2018. Approximating optimal social choice under metric preferences. *Artific. Intell.* 264 (2018), 27–51.
- [5] Elliot Anshelevich, Onkar Bhardwaj, and John Postl. 2015. Approximating optimal social choice under metric preferences. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. AAAI Press, 777–783.

- [6] Elliot Anshelevich and John Postl. 2017. Randomized social choice functions under metric preferences. *J. Artif. Intell. Res.* 58 (2017), 797–827.
- [7] Elliot Anshelevich and Shreyas Sekar. 2016. Blind, greedy, and random: Algorithms for matching and clustering using only ordinal information. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*. 383–389.
- [8] Elliot Anshelevich and Shreyas Sekar. 2016. Truthful mechanisms for matching and clustering in an ordinal world. In *Proceedings of the International Conference on Web and Internet Economics (WINE'16)*. Springer, 265–278.
- [9] Elliot Anshelevich and Wennan Zhu. 2019. Tradeoffs between information and ordinal approximation for bipartite matching. *Theory Comput. Syst.* 63, 7 (2019), 1499–1530.
- [10] Gerdus Benade, Swaprava Nath, Ariel D. Procaccia, and Nisarg Shah. 2017. Preference elicitation for participatory budgeting. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. 376–382.
- [11] Gerdus Benade, Ariel D. Procaccia, and Mingda Qiao. 2019. Low-distortion social welfare functions. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. 1788–1795.
- [12] Anand Bhalgat, Deeparnab Chakrabarty, and Sanjeev Khanna. 2011. Social welfare in one-sided matching markets without money. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 87–98.
- [13] Umang Bhaskar, Varsha Dani, and Abheek Ghosh. 2018. Truthful and near-optimal mechanisms for welfare maximization in multi-winner elections. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. 925–932.
- [14] Anna Bogomolnaia and Hervé Moulin. 2004. Random matching under dichotomous preferences. *Econometrica* 72, 1 (2004), 257–279.
- [15] Allan Borodin, Omer Lev, Nisarg Shah, and Tyrone Strangway. 2019. Primarily about Primaries. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. 1804–1811.
- [16] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D. Procaccia, and Or Sheffet. 2015. Optimal social choice functions: A utilitarian view. *Artif. Intell.* 227 (2015), 190–213.
- [17] Jarosław Byrka, Thomas Pensch, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. 2014. An improved approximation for k-median, and positive correlation in budgeted optimization. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'14)*. Society for Industrial and Applied Mathematics, 737–756.
- [18] Ioannis Caragiannis, Aris Filos-Ratsikas, Søren Kristoffer Stiil Frederiksen, Kristoffer Arnsfelt Hansen, and Zihan Tan. 2016. Truthful facility assignment with resource augmentation: An exact analysis of serial dictatorship. In *Proceedings of the International Conference on Web and Internet Economics (WINE'16)*. Springer, 236–250.
- [19] Ioannis Caragiannis, Swaprava Nath, Ariel D. Procaccia, and Nisarg Shah. 2017. Subset selection via implicit utilitarian voting. *J. Artif. Intell. Res.* 58 (2017), 123–152.
- [20] Yu Cheng, Shaddin Dughmi, and David Kempe. 2017. Of the people: Voting is more effective with representative candidates. In *Proceedings of the ACM Conference on Economics and Computation (EC'17)*. ACM, 305–322.
- [21] Yu Cheng, Shaddin Dughmi, and David Kempe. 2018. On the distortion of voting with multiple representative candidates. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. 973–980.
- [22] George Christodoulou, Aris Filos-Ratsikas, Søren Kristoffer Stiil Frederiksen, Paul W. Goldberg, Jie Zhang, and Jinshan Zhang. 2016. Social welfare in one-sided matching mechanisms. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS'16)*. Springer, 30–50.
- [23] Ehsan Emamjomeh-Zadeh and David Kempe. 2018. Adaptive hierarchical clustering using ordinal queries. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'18)*. Society for Industrial and Applied Mathematics, 415–429.
- [24] James M. Enelow and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge University Press, New York, NY.
- [25] Brandon Fain, Ashish Goel, Kamesh Munagala, and Nina Prabhu. 2019. Random dictators with a random referee: Constant sample complexity mechanisms for social choice. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 1893–1900.
- [26] Brandon Fain, Ashish Goel, Kamesh Munagala, and Sukolsak Sakshuwong. 2017. Sequential deliberation for social choice. In *Proceedings of the International Conference on Web and Internet Economics (WINE'17)*. Springer, 177–190.
- [27] Reza Zanjirani Farahani and Masoud Hekmatfar. 2009. *Facility Location: Concepts, Models, Algorithms and Case Studies*. Springer.
- [28] Michal Feldman, Amos Fiat, and Iddan Golomb. 2016. On voting and facility location. In *Proceedings of the ACM Conference on Economics and Computation (EC'16)*. ACM, 269–286.
- [29] Aris Filos-Ratsikas, Søren Kristoffer Stiil Frederiksen, and Jie Zhang. 2014. Social welfare in one-sided matchings: Random priority and beyond. In *Proceedings of the International Symposium on Algorithmic Game Theory (SAGT'14)*. Springer, 1–12.
- [30] Aris Filos-Ratsikas, Evi Micha, and Alexandros A. Voudouris. 2019. The distortion of distributed voting. In *Proceedings of the International Symposium on Algorithmic Game Theory (SAGT'19)*. Springer, 312–325.

- [31] Mohammad Ghodsi, Mohamad Latifian, and Masoud Seddighin. 2019. On the distortion value of the elections with abstention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 1981–1988.
- [32] Ashish Goel, Reyna Hulett, and Anilesh K. Krishnaswamy. 2018. Relating metric distortion and fairness of social choice rules. In *Proceedings of the 13th Workshop on the Economics of Networks, Systems and Computation (NetEcon'18)*.
- [33] Ashish Goel, Anilesh K. Krishnaswamy, and Kamesh Munagala. 2017. Metric distortion of social choice rules: Lower bounds and fairness properties. In *Proceedings of the ACM Conference on Economics and Computation (EC'17)*. ACM, 287–304.
- [34] Ashish Goel, Anilesh K. Krishnaswamy, Sukolsak Sakshuwong, and Tanja Aitamurto. 2019. Knapsack voting for participatory budgeting. *ACM Trans. Econ. Comput.* 7, 2 (2019), 8.
- [35] Stephen Gross, Elliot Anshelevich, and Lirong Xia. 2017. Vote until two of you agree: Mechanisms with small distortion and sample complexity. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. 544–550.
- [36] Dorit S. Hochbaum and David B. Shmoys. 1985. A best possible heuristic for the k-center problem. *Math. Oper. Res.* 10, 2 (1985), 180–184.
- [37] Martin Hoefer and Bojana Kodric. 2017. Combinatorial secretary problems with ordinal information. In *Proceedings of the 44th International Colloquium on Automata, Languages, and Programming (ICALP'17)*. 133:1–133:14.
- [38] David Kempe. 2020. An analysis framework for metric voting based on LP duality. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*.
- [39] David Kempe. 2020. Communication, distortion, and randomness in metric voting. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*.
- [40] Shi Li. 2011. A 1.488 approximation algorithm for the uncapacitated facility location problem. In *Proceedings of the International Colloquium on Automata, Languages, and Programming (ICALP'11)*. Springer, 77–88.
- [41] Debmalya Mandal, Ariel D. Procaccia, Nisarg Shah, and David Woodruff. 2019. Efficient and thrifty voting by any means necessary. In *Advances in Neural Information Processing Systems*. MIT Press, 7178–7189.
- [42] Kamesh Munagala and Kangning Wang. 2019. Improved metric distortion for deterministic social choice rules. In *Proceedings of the ACM Conference on Economics and Computation (EC'19)*. ACM, 245–262.
- [43] Grzegorz Pierczyński and Piotr Skowron. 2019. Approval-based elections and distortion of voting rules. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19)*. 543–549.
- [44] Ariel D. Procaccia and Jeffrey S. Rosenschein. 2006. The distortion of cardinal preferences in voting. In *Proceedings of the International Workshop on Cooperative Information Agents (CIA'06)*. Springer, 317–331.
- [45] Piotr Krzysztof Skowron and Edith Elkind. 2017. Social choice under metric preferences: Scoring rules and STV. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. 706–712.

Received April 2019; revised December 2019; accepted May 2020