A Few Queries Go a Long Way: Information-Distortion Tradeoffs in Matching

Georgios Amanatidis,[1,2] Georgios Birmpas,[3] Aris Filos-Ratsikas,[4] Alexandros A. Voudouris[5]

[1]Department of Mathematical Sciences, University of Essex
[2]ILLC, University of Amsterdam
[3]Department of Computer, Control and Management Engineering, Sapienza University of Rome
[4]Department of Computer Science, University of Liverpool
[5]School of Computer Science and Electronic Engineering, University of Essex
georgios.amanatidis@essex.ac.uk, george.birbas@diag.uniroma1.it, aris.filos-ratsikas@liverpool.ac.uk,
alexandros.voudouris@essex.ac.uk

Abstract

We consider the *one-sided matching* problem, where n agents have preferences over n items, and these preferences are induced by underlying cardinal valuation functions. The goal is to match every agent to a single item so as to maximize the *social welfare*. Most of the related literature, however, assumes that the values of the agents are not a priori known, and only access to the ordinal preferences of the agents over the items is provided. Consequently, this incomplete information leads to loss of efficiency, which is measured by the notion of *distortion*. In this paper, we further assume that the agents can answer a small number of *queries*, allowing us partial access to their values. We study the interplay between elicited cardinal information (measured by the number of queries per agent) and distortion for one-sided matching, as well as a wide range of well-studied related problems. Qualitatively, our results show that with a limited number of queries, it is possible to obtain significant improvements over the classic setting, where only access to ordinal information is given.

1 Introduction

In the *one-sided matching* problem (often referred to as the *house allocation* problem), n agents have *preferences* over a set of n items, and the goal is to find an allocation in which every agent receives a single item, while maximizing some objective. Typically, as well as in this paper, this objective is the (utilitarian) social welfare, i.e., the total utility of the agents. Since its introduction by Hylland and Zeckhauser (1979), this has been one of the most fundamental problems in the literature of economics (e.g., see (Bogomolnaia and Moulin 2001; Svensson 1999)), and has also been extensively studied in computational social choice (e.g., see (Klaus, Manlove, and Rossi 2016)).

The classic work on the problem (including Hylland and Zeckhauser's seminal paper) assumes that the preferences of the agents are captured by cardinal valuation functions, assigning numerical values to the different items; these can be interpreted as their *von Neuman-Morgenstern utilities* (Von Neumann and Morgenstern 1947). From a more algorithmic viewpoint, one can envision a weighted complete bipartite graph (with agents and items forming the two sides of the partition), where the weights of the edges are given by these values. Crucially, most of the related literature assumes that the designer only has access to the preference rankings of the agents over the items (i.e., the *ordinal preferences*) induced by the underlying values, but not to the values themselves.[1] This is motivated by the fact that it is fairly standard to ask the agents to simply order the items, while it is arguably much more demanding to require them to specify exact numerical values for all of them.

This begs the following natural question: *What is the effect of this limited information on the goals of the algorithm designer?* In 2006, Procaccia and Rosenschein defined the notion of *distortion* to measure precisely this effect, when the goal is to maximize the social welfare. Their original research agenda was put forward for settings in general social choice (also referred to as *voting*), but has since then flourished to capture several different scenarios, including the one-sided matching problem. For the latter problem, Filos-Ratsikas, Frederiksen, and Zhang (2014), showed that the best possible distortion achieved by any ordinal algorithm is $\Theta(\sqrt{n})$, even if one allows randomization, and even if the valuations are normalized. For deterministic algorithms, the corresponding bound is $\Theta(n^2)$ (Theorem 1).

While the aforementioned bounds establish a stark impossibility when one has access only to ordinal information, they do not rule out the prospect of good approximations when it is possible to elicit *some* cardinal information. Indeed, the cognitive burden of eliciting cardinal values in the literature has mostly been considered in the two extremes; either full cardinal information or not at all. Conceivably though, if the agents needed to come up with only *a few* cardinal values, the elicitation process would not be very demanding, while it could potentially have wondrous effects on the social welfare. This approach was advocated recently by Amanatidis et al. (2020b), who proposed to study the *tradeoffs* between the number of cardinal *value queries* per agent and distortion. For the general social choice setting of Procaccia and Rosenschein (2006), Amanatidis et al. (2020b) actually showed that with a limited number of such queries, one can significantly improve upon the existing strong impossibilities (Boutilier et al. 2015; Caragiannis et al. 2017).

[1] The pseudo-market mechanism of Hylland and Zeckhauser (1979) is a notable exception to this.

Motivated by the success of this approach for general social choice settings, we extend this research agenda and aim to answer the following question for the one-sided matching problem:

What are the best possible information-distortion tradeoffs in one-sided matching? Can we achieve significant improvements over the case of only ordinal preferences, by making only a few cardinal value queries per agent?

1.1 Our Contribution

We consider the one-sided matching problem with the goal of maximizing the social welfare under limited information. We adopt the standard assumption in the related literature that the agents provide as input their ordinal preferences over the items, and that these are induced by their cardinal valuation functions. Following the agenda put forward by Amanatidis et al. (2020b), we also assume implicit access to the numerical values of the agents via *value queries*; we may ask for an agent i and an item j, and obtain the agent's value, $v_i(j)$, for that item.

We measure the performance of an algorithm by the standard notion of *distortion*, and our goal is to explore the tradeoffs between distortion and the number of queries we need per agent. As the two extremes, we note that if we use n queries per agent, we recover the complete cardinal valuation profile and thus the distortion is 1, whereas if we use 0 queries, i.e., we use only the ordinal information, the best possible distortion is $\Theta(n^2)$ (see Theorem 1). The latter bound holds even if we consider valuation functions that satisfy the *unit-sum* normalization, i.e., the sum of the values of each agent for all the items is 1. As we mentioned earlier, even when allowing randomization, the best possible distortion is still quite large ($\Theta(\sqrt{n})$ (Filos-Ratsikas, Frederiksen, and Zhang 2014)) without employing any value queries. In this work, we only consider deterministic algorithms, and leave the study of randomized algorithms for future work.

We provide the following results:

- In Section 3, we present an algorithm parametrized by λ, which achieves distortion $O(n^{1/(\lambda+1)})$ by making $O(\lambda \log n)$ queries per agent. In particular, by setting $\lambda = O(1)$ and $\lambda = O(\log n)$ we achieve respectively
  - distortion $O(\sqrt{n})$ using $O(\log n)$ queries per agent;
  - *constant* distortion using $O(\log^2 n)$ queries per agent.

  The algorithm is inspired by a conceptually similar idea presented by Amanatidis et al. (2020b) for the social choice setting. In Section 6 we adapt our algorithm to provide analogous information-distortion tradeoffs for a wide range of well-studied optimization problems, including *two-sided matching*, *general graph matching* and the *clearing problem for kidney exchange*.
- In Section 4 we show a lower bound of $\Omega(n^{1/k}/k)$ on the distortion of any algorithm that makes k queries per agent. An immediate consequence is that it is impossible to achieve constant distortion without asking almost $\log n$ queries! When k is a constant, it is possible to show that our construction is tight. Furthermore, using a construction which exploits the same ordinal but different cardinal information, we can show that even under the stronger assumption of unit-sum normalization, the distortion cannot be better than $\Omega\left(n^{1/(k+1)}/k\right)$ with k queries per agent.
- In Section 5 we present our main algorithmic result for unit-sum valuations, namely a novel algorithm which achieves distortion $O(n^{2/3}\sqrt{\log n})$ using only *two* queries per agent.

Our results are summarized in Figure 1. We note that our upper bounds are robust to "errors" in the responses to the queries. When the reported values are within a multiplicative factor r from the true values, this error parameter r enters multiplicatively into the bounds. In the reasonable case where r is constant, our bounds (which are asymptotic) are unaffected. For the sake of readability, we state our results without assuming any such errors.

1.2 Related Work

One-sided matching in the context of agents with preferences over items was introduced by Hylland and Zeckhauser (1979). The classic literature in economics (e.g., see (Bogomolnaia and Moulin 2001; Svensson 1999)) is mostly concerned with axiomatic properties, and has proposed several solutions and impossibilities; see the surveys of Sönmez and Ünver (2011) and Abdulkadiroglu and Sönmez (2013) for more information.

The effects of limited information on the social welfare objective were studied most notably in the work of Filos-Ratsikas, Frederiksen, and Zhang (2014) mentioned earlier. Further, Anshelevich and Sekar (2016a), Anshelevich and Sekar (2016b), Anshelevich and Zhu (2017), and Abramowitz and Anshelevich (2018) studied related settings on graphs, and showed distortion bounds for matching problems and their generalizations. A crucial difference from our work is that these papers consider edges with *symmetric* weights. In contrast, in our case the weights that are induced by the values of the agents are *asymmetric*, which makes the results markedly different; see (Anshelevich, Das, and Naamad 2013) for a more detailed discussion. For general social choice settings (i.e., voting), the distortion of ordinal algorithms has been studied in a long list of papers, e.g., see (Anshelevich and Postl 2017; Anshelevich et al. 2018; Boutilier et al. 2015) and references therein.

The approach of enhancing the input of algorithms by equipping them with cardinal queries that we adopt in this paper was first suggested by Amanatidis et al. (2020b). Some other works (Abramowitz, Anshelevich, and Zhu 2019; Benade et al. 2017; Bhaskar, Dani, and Ghosh 2018) have considered related but different models in which the designer has access to some cardinal information on top of the ordinal preferences. In a recent orthogonal approach, Mandal et al. (2019) and Mandal, Shah, and Woodruff (2020) considered the communication complexity of voting algorithms and studied the tradeoffs between the distortion and the number of bits of information elicited from the agents.

	ordinal info, 0 queries	2 queries	k queries ($k = \Theta(1)$)	k queries ($k = \omega(1)$)	$O(\lambda \log n)$ queries	$O(\log^2 n)$ queries	full info, n queries
unit-sum valuations	$\Theta\left(n^2\right)$	$\tilde{O}\left(n^{2/3}\right)$ $\Omega\left(n^{1/3}\right)$	$\Omega\left(n^{1/(k+1)}\right)$	$\Omega\left(\frac{n^{1/(k+1)}}{k}\right)$			
unrestricted valuations	unbounded	$\Omega\left(\sqrt{n}\right)$	$\Omega\left(n^{1/k}\right)$	$\Omega\left(\frac{n^{1/k}}{k}\right)$	$O\left(n^{1/(\lambda+1)}\right)$	$O(1)$	1

Figure 1: An overview of our results on the number of queries per agent. The upper bounds that hold for unrestricted valuations also hold for unit-sum valuations. All of the lower bounds hold even when all agents have the same ordinal preferences.

2 Model Definition

We consider the one-sided matching problem, where there is a set of agents N and a set of items A, such that $|N| = |A| = n$. Each agent $i \in N$ has a *valuation function* $v_i : A \to \mathbb{R}_{\geq 0}$ indicating the agent's value for each item; that is $v_i(j)$ is the value of agent $i \in N$ for item $j \in A$. The valuation functions we consider are either

- *unrestricted*, in which case the values for the items can be any non-negative real numbers, or
- *unit-sum*, in which case the sum of values of each agent i for all items is 1: $\sum_{j \in A} v_i(j) = 1$.

We denote by $\mathbf{v} = (v_i)_{i \in N}$ the *(cardinal) valuation profile* of the agents. Let $Y = (y_i)_{i \in N}$ be a *matching* according to which each agent $i \in N$ is matched to exactly one item $y_i \in A$, such that $y_i \neq y_{i'}$ for every $i \neq i'$. Given a profile $\mathbf{v}$, the *social welfare* of Y, $\text{SW}(Y|\mathbf{v})$, is the total value of the agents for the items they are matched to according to Y:

$$\text{SW}(Y|\mathbf{v}) = \sum_{i \in N} v_i(y_i)\,.$$

By $\mathcal{M}$ we denote the set of all perfect matchings on our instance. Our goal is to compute a matching $X(\mathbf{v}) = (x_i)_{i \in N}$ with maximum social welfare, i.e.,

$$X(\mathbf{v}) \in \arg\max_{Y \in \mathcal{M}} \text{SW}(Y|\mathbf{v})\,.$$

In case the valuation functions of the agents are known, then computing $X(\mathbf{v})$ can be done efficiently, e.g., via the Hungarian method (Kuhn 1956). However, our setting is a bit more restrictive. The exact valuation functions of the agents are their private information, and they can instead report orderings over the items, which are consistent with their valuations. In particular, every agent i reports a ranking of the items $\succ_i$ such that $a \succ_i b$ if and only if $v_i(a) \geq v_i(b)$ for all $a, b \in A$. Given a valuation profile $\mathbf{v}$, we denote by $\succ_{\mathbf{v}} = (\succ_i)_{i \in N}$ the *ordinal profile* induced by $\mathbf{v}$; observe that different valuation profiles may induce the same ordinal profile. On top of the ordinal preferences of the agents, we can obtain partial access to the valuation profile, by making a number of value queries per agent. In particular, a *value query* takes as input an agent $i \in N$ and an item $j \in A$, and returns the value $v_i(j)$ of agent i for item j. This leads us to the following definition of a deterministic algorithm in our setting.

Definition 1. A *matching algorithm* $\mathcal{A}_k$ takes as input an ordinal profile $(\succ)_{i \in N}$, makes $k \leq n$ value queries per agent and, using $(\succ)_{i \in N}$ as well as the answers to the queries, it computes a matching $\mathcal{A}_k(\succ) \in \mathcal{M}$. If $k = 0$, $\mathcal{A}$ is an *ordinal* algorithm, whereas if $k = n$, $\mathcal{A}$ is a *cardinal* algorithm.

As already mentioned, we can efficiently compute the optimal matching using a cardinal algorithm. However, if an algorithm is allowed to make a limited number $k < n$ of queries per agent, the computed matching might not be optimal. The question then is how well does such an algorithm *approximate* the optimal social welfare of any matching. Approximation here is captured by the notion of *distortion*.

Definition 2. The *distortion* $\text{dist}(\mathcal{A}_k)$ of an algorithm $\mathcal{A}_k$ is the worst-case ratio (over all possible valuation profiles $\mathbf{v}$) between the social welfare of an optimal matching $X(\mathbf{v})$ and the social welfare of the matching computed by $\mathcal{A}_k$:

$$\text{dist}(\mathcal{A}_k) = \sup_{\mathbf{v}} \frac{\text{SW}(X(\mathbf{v})|\mathbf{v})}{\text{SW}(\mathcal{A}_k(\succ_{\mathbf{v}})|\mathbf{v})}\,.$$

Warm-up: Ordinal Algorithms. Before we proceed with our more technical results on tradeoffs between information and distortion, we consider the case of ordinal algorithms. When the valuation functions of the agents are unrestricted, the distortion of any ordinal algorithm is unbounded. To see this, consider any instance that contains two agents who agree on which the most valuable item is. Since only one of them can be matched to this item, it might be the case that the other agent has an arbitrarily large value for it, leading to unbounded distortion. Even for the more restrictive case of unit-sum valuations, however, the distortion of ordinal algorithms can be quite large.

Theorem 1. *For unit-sum valuation functions, the distortion of the* best *ordinal matching algorithm is* $\Theta(n^2)$.

Due to space constraints, the proof of the theorem, as well as several other proofs are deferred to the full version of this work (Amanatidis et al. 2020a).

3 Guarantees for Unconstrained Valuations

Here we present λ-THRESHOLDSTEPFUNCTION (λ-TSF), an algorithm that works for any valuation functions. At a high level, for each agent, we do the following. We first query the agent's value for her highest ranked item. Then, we partition the items into $\lambda + 1$ sets, so that the agent's value for all the items in a set is lower-bounded by a carefully selected quantity related to the agent's top value. Based on this partition, we then define a new *simulated valuation function* for the agent, where the value of an item is equal to the lower bound that corresponds to the set the item belongs to. Finally, we compute a maximum weight matching with respect to the simulated valuation functions. Formally:

> λ-THRESHOLDSTEPFUNCTION (λ-TSF)
> Let $\alpha_\ell = n^{-\ell/(\lambda+1)}$ for $\ell \in \{0, ..., \lambda\}$.
> For every agent $i \in N$:
> - Query i for her top-ranked item j_i^*; let v_i^* be this value.
> - Let $Q_{i,0} = \{j_i^*\}$ and $\tilde{v}_i(j_i^*) = \alpha_0 \cdot v_i^* = v_i^*$.
> - For every $\ell \in \{1, ..., \lambda\}$, using binary search, compute
>
> $$Q_{i,\ell} = \{j \in A : v_i(j) \in [\alpha_\ell \cdot v_i^*, \alpha_{\ell-1} \cdot v_i^*)\}$$
>
> and let $\tilde{v}_i(j) = \alpha_\ell \cdot v_i^*$ for every $j \in Q_{i,\ell}$.
> - Let $Q_i = \bigcup_{\ell=0}^{\lambda} Q_{i,\ell}$ and set $\tilde{v}_i(j) = 0$ for $j \in A \setminus Q_i$.
>
> Return a matching $Y \in \arg\max_{Z \in \mathcal{M}} \text{SW}(Z|\tilde{\mathbf{v}})$.

For each i and ℓ, in order to find i's least preferred item that she values at least $\alpha_\ell v_i^*$, in the third bullet we run a standard binary search on $\succ_i$. It is known that each such binary search requires $1 + \log_2 n$ queries. The next theorem follows by the more general Theorem 7, which is stated in Section 6 and applies to a number of well-known graph problems.

Theorem 2. *λ-TSF makes $1 + \lambda + \lambda \log n$ queries per agent and achieves a distortion of $2n^{1/(\lambda+1)}$.*

By appropriately setting the value of λ, we obtain several tradeoffs between the distortion and the number of queries per agent. In particular, we have the following statement.

Corollary 1. *We can achieve*

- *distortion $O(n)$ by making* one *query per agent;*
- *distortion $O(n^{1/k})$ for any constant integer k by making $O(\log n)$ queries per agent;*
- *distortion $O(1)$ by making $O(\log^2 n)$ queries per agent.*

4 Lower Bounds

In this section we show unconditional lower bounds for algorithms for one-sided matching which are allowed to make at most $k \geq 1$ queries per agent. We present a generic matching instance which can be fine-tuned to yield lower bounds for both unrestricted and unit-sum valuation functions. Let $\mathcal{V}$ denote any of these two classes of valuation functions.

Let $\delta_\mathcal{V}(k) \leq 1/k$ be a function of k, and $\varepsilon \in (0, 1/2)$ be some constant. We want to define an instance in which the n items are partitioned into $k+2$ sets $A_1, ..., A_{k+1}, B = A \setminus \left(\bigcup_{\ell \in [k+1]} A_\ell\right)$ such that

$$|A_\ell| = \varepsilon \cdot n^{(\ell-1)\delta_\mathcal{V}(k)}.$$

We assume that n is large enough so that $n > 2\sum_{\ell=1}^{k+1} |A_\ell|$. Using the notation $\langle T \rangle$ for some arbitrary fixed ranking of the elements of set T (which is common for all agents), we define the ordinal preference of every agent $i \in N$ to be

$$\langle A_1 \rangle \succ_i \langle A_2 \rangle \succ_i ... \succ_i \langle A_k \rangle \succ_i \langle A_{k+1} \rangle \succ_i \langle B \rangle.$$

We reveal the following information, depending on the queries of the algorithm:

- For every $\ell \in \{1, \dots, k+1\}$, any query for some item in A_ℓ reveals a value of $|A_\ell|^{-1} \cdot n^{-\delta_\mathcal{V}(k)}$;
- Every query for some item in B reveals a value of 0.

Next, we define two types of conditional valuation functions that an agent i may have, depending on the behavior of the algorithm. These functions have to be consistent to the information that is revealed by the queries of the algorithm. Let $\xi \in (0, 1]$ be some constant.

(T1) If there exists $r \in \{1, \dots, k+1\}$, such that i is *not* queried for any item in A_r and she *does not* get an item from A_r either, then i's values are

- at least $\xi \cdot |A_{r-1}|^{-1} \cdot n^{-\delta_\mathcal{V}(k)}$ for each item in A_r if $r \geq 2$;
- at least ξ for the item in A_1 if $r = 1$;
- $|A_\ell|^{-1} \cdot n^{-\delta_\mathcal{V}(k)}$ for every item in A_ℓ, $\ell \in \{1, \dots, k+1\} \setminus \{r\}$;
- 0 for every item in B.

(T2) If i is queried for some item in k different sets out of $A_1, \dots, A_{k+1}$, then her values are

- $|A_\ell|^{-1} \cdot n^{-\delta_\mathcal{V}(k)}$ for every item in A_ℓ, $\ell \in \{1, \dots, k+1\}$;
- at most $|A_{k+1}|^{-1} \cdot n^{-\delta_\mathcal{V}(k)}$ for every item in B.

Observe that the conditions specified in (T1) and (T2) capture all possible cases about the queries of the algorithm and the possible assignments of the items to the agents.

Theorem 3. *Let $\mathcal{V}$ be the class of unrestricted or unit-sum valuation functions. If there exists a function $\delta_\mathcal{V}(k) \leq 1/k$ such that it is possible to define valuation functions in $\mathcal{V}$ of types* (T1) *and* (T2)*, the distortion of any matching algorithm which makes k queries per agent is $\Omega\big(\frac{1}{k} \cdot n^{\delta_\mathcal{V}(k)}\big)$.*

Theorem 3 is actually quite powerful and allows us to prove lower bounds for both unrestricted and unit-sum valuation functions. In particular, it reduces the problem to finding the largest possible $\delta_\mathcal{V}(k) \leq 1/k$, such that valuation functions in $\mathcal{V}$ of types (T1) and (T2) can be defined.

Theorem 4. *For unconstrained valuation functions, the distortion of any matching algorithm which makes k queries per agent is $\Omega\big(\frac{1}{k} \cdot n^{1/k}\big)$.*

Before we state the corresponding bound for unit-sum valuations, we remark that, for any constant k, the bound of Theorem 4 is tight with respect to this particular construction. That is, we cannot hope to prove stronger lower bounds using this class of instances. For unit-sum valuations, we have the following bound.

Theorem 5. *For unit-sum valuation functions, the distortion of any matching algorithm which makes $k \leq (1-\xi)n^{1/(k+1)}$ queries per agent is $\Omega\big(\frac{1}{k} \cdot n^{1/(k+1)}\big)$.*

By appropriately setting the value of k in Theorems 4 and 5, we establish that it is impossible to achieve constant distortion without an almost logarithmic number of queries.

Corollary 2. *Any matching algorithm allowed to make $o\big(\frac{\log n}{\log\log n}\big)$ queries per agent has distortion $\omega(\log\log n)$.*

5 Two Queries for Unit-sum Valuations

In this section, we present the FIRSTPOSITIONADAPTIVE algorithm (FPA), which makes at most two queries per agent and achieves a distortion of $O(n^{2/3}\sqrt{\log n})$, when the valuation functions are unit-sum. First, we query each agent for their most-preferred item. Then, depending on whether the *maximum revealed value* by these queries is at least $n^{-1/3}$, we query the agents for items that are parts of "large enough" partial matchings. Otherwise, we query everyone at a specific position, and define *simulated values* based on the answers to these queries, ensuring that these values are lower bounds on the corresponding true values. Clearly, the simulated valuation functions are not necessarily unit-sum.

For the sake of presentation, we assume that n is a perfect cube, i.e., $n = \alpha^3$ for some $\alpha \in \mathbb{N}$. Formally:

FIRSTPOSITIONADAPTIVE (FPA)

All agents are initially *active*.

For every agent i, query i for her top item j_i^*; let v_i^* be its value.

If $\max_{i\in N} v_i^* \geq n^{-1/3}$, then:

- For every $\ell \in [n]$, *while* there exists a partial matching Y_p of size $|Y_p| \geq n^{1/3}/\sqrt{\log n}$ consisting of active agents i matched to items y_i such that agent i ranks item y_i at some position $\ell' \leq \ell$, query every i in Y_p for item y_i and make these agents *inactive*. If necessary, break ties arbitrarily.
- Output a matching Y that maximizes the social welfare, according to the revealed values due to the above queries.

Else (i.e., $\max_{i\in N} v_i^* < n^{-1/3}$):

- For every agent i, query i for the item she ranks at position $n^{1/3}+1$; let u_i be this value.
- For every agent i, define the simulated valuation function $\tilde{v}_i$:
  - $\tilde{v}_i(j_i{}^*) = v_i^*$;
  - $\tilde{v}_i(j) = u_i$ for every item j that i ranks at position $\ell \in \{2, \dots, n^{1/3}+1\}$;
  - $\tilde{v}_i(j) = 0$ for every item j that i ranks at position $\ell \in \{n^{1/3}+2, \dots, n\}$.
- For every agent i such that $u_i < \frac{1}{2}n^{-1}$, modify $\tilde{v}_i$ so that:
  - $\tilde{v}_i(j) = \frac{1}{3}n^{-1/3}$ for every item j that i ranks at position $\ell \in \{2, \dots, \frac{1}{4}n^{1/3}\}$.
- Output a matching $Y \in \arg\max_Z \text{SW}(Z|\tilde{\mathbf{v}})$.

Theorem 6. *For unit-sum valuation functions, the distortion of FPA is $O(n^{2/3}\sqrt{\log n})$.*

Proof. Let $\mathbf{v}$ be a valuation profile. Denote by Y the output of the algorithm when given as input the ordinal profile $\succ_{\mathbf{v}}$, and by $X = (x_i)_{i\in N}$ an optimal matching for $\mathbf{v}$. We consider two main cases, depending on the value $\max_{i\in N} v_i^*$ that the algorithm learns with the first query.

Case 1: $\max_{i\in N} v_i^* \geq n^{-1/3}$

The algorithm makes a second query to an agent i for some item j only if the pair (i, j) is part of a partial matching of size at least $n^{1/3}$, involving only active agents, i.e., agents who have not been included in such a partial matching in any previous step. Let $Y_1, \dots, Y_\lambda$ be all the partial matchings considered throughout the execution of the algorithm. By definition, each such partial matching contains at least $n^{1/3}/\sqrt{\log n}$ agents and an agent is contained in at most one of these matchings. Thus, it holds that $\lambda < n^{2/3}\sqrt{\log n}$.

We partition the agents into two sets. The set H of agents i that are queried only for items they rank *at least as high* as the item x_i they receive in the optimal matching X. Some agents in H are possibly queried twice for their best item. The set L of agents i that are queried for an item they rank *lower* than x_i or are queried only *once*. We can write the social welfare of X as

$$\text{SW}(X|\mathbf{v}) = \sum_{i\in H} v_i(x_i) + \sum_{i\in L} v_i(x_i)\,.$$

We will bound each term on the right-hand side separately. For the first term, we have:

$$\sum_{i\in H} v_i(x_i) \leq \sum_{i\in H} v_i(y_i) \leq \sum_{t=1}^{\lambda}\sum_{i\in Y_t} v_i(y_i) \leq \lambda \max_t \sum_{i\in Y_t} v_i(y_i) < n^{2/3}\sqrt{\log n}\cdot \text{SW}(Y|\mathbf{v}).$$

The first inequality holds because $y_i \succcurlyeq_i x_i$ for every $i \in H$. The second inequality holds because the agents in H are queried only if they are included in one of the partial matchings $Y_1, \dots, Y_\lambda$. The last inequality follows from the bound on λ established above, and the fact that $\max_t \sum_{i\in Y_t} v_i(y_i)$ is trivially upper bounded by the social welfare of Y.

To bound the second term, let $Y^{(\ell)}$ be the restriction of Y containing only the agents $i \in L$ for whom x_i is at position ℓ. It holds that $|Y^{(\ell)}| < n^{1/3}/\sqrt{\log n}$, or else the algorithm would have queried the agents in $Y^{(\ell)}$ for their optimal items, contradicting their membership in L. So, we get that

$$\sum_{i\in L} v_i(x_i) = \sum_{\ell=1}^{n}\sum_{i\in Y^{(\ell)}} v_i(x_i) < \sum_{\ell=1}^{n} n^{1/3}\sqrt{\log n}\,\frac{1}{\ell} < \frac{n^{1/3}}{\sqrt{\log n}} 2\log n = 2n^{1/3}\sqrt{\log n}\,,$$

where the first inequality follows from the unit-sum normalization; in particular, any agent's value for an item at position ℓ is at most $1/\ell$. The second inequality is a simple bound on the harmonic numbers: $\sum_{i=1}^{n} i^{-1} < 2\log_2 n$, for $n \geq 2$.

Further, since $\max_{i\in N} v_i^* \geq n^{-1/3}$, we have that $\text{SW}(Y|\mathbf{v}) \geq n^{-1/3}$. Thus,

$$\sum_{i\in L} v_i(x_i) \leq 2n^{2/3}\sqrt{\log n}\cdot \text{SW}(Y|\mathbf{v})\,.$$

Putting everything together, the distortion of the algorithm in this case is at most $2n^{2/3}\sqrt{\log n}$.

Case 2: $\max_{i\in N} v_i^* < n^{-1/3}$

We partition the set of agents into two sets, depending on whether their value for the item they rank at position $n^{1/3}+1$ is at most $\frac{1}{2}n^{-1}$. In particular, let $R = \{i \in N : u_i < \frac{1}{2}n^{-1}\}$. We can write the optimal social welfare of X as

$$\text{SW}(X|\mathbf{v}) = \sum_{i\in R} v_i(x_i) + \sum_{i\in N\setminus R} v_i(x_i)\,.$$

We will bound each term separately. For the first term, since $\max_{i\in N} v_i^* < n^{-1/3}$, we clearly have that

$$\sum_{i\in R} v_i(x_i) \le \max_{i\in N} v_i^*|R| \le n^{-1/3}|R|\,.$$

Consider an arbitrary agent $i \in R$ and denote by $j_{i,\ell}$ the item she ranks at position ℓ; hence, $j_i^* = j_{i,1}$. We will first show that $v_i\big(j_{i,\frac{1}{4}n^{1/3}}\big) \ge \frac{1}{3}\cdot n^{-1/3} = \tilde{v}_i\big(j_{i,\frac{1}{4}n^{1/3}}\big)$. Since $u_i = v_i\big(j_{i,n^{1/3}+1}\big) < \frac{1}{2}n^{-1}$, we have that $\sum_{\ell=n^{1/3}+1}^{n} v_i(j_{i,\ell}) \le (n - n^{1/3} - 1)u_i < \frac{1}{2}$, and thus, by the unit-sum normalization assumption, we also have that $\sum_{\ell=1}^{n^{1/3}} v_i(j_{i,\ell}) \ge \frac{1}{2}$. Since $v_i(j_{i,\ell}) \le v_i(j_{i,1}) < n^{-1/3}$ for every $\ell \in \{1,\dots,\frac{1}{4}n^{1/3}-1\}$ and $v_i\big(j_{i,\frac{1}{4}n^{1/3}}\big) \ge v_i(j_{i,\ell})$ for every $\ell \in \{\frac{1}{4}n^{1/3},\dots,n^{1/3}\}$, we obtain

$$\begin{aligned} v_i\big(j_{i,\frac{1}{4}n^{1/3}}\big) &\ge \frac{\frac{1}{2} - \big(\frac{1}{4}n^{1/3}-1\big)n^{-1/3}}{\frac{3}{4}n^{1/3}} \\ &\ge \frac{1}{3}n^{-1/3} = \tilde{v}_i\big(j_{i,\frac{1}{4}n^{1/3}}\big)\,. \end{aligned}$$

So, all the agents in R have value at least $\frac{1}{3}n^{-1/3}$ for the items they rank at positions up to $\frac{1}{4}n^{1/3}$. This implies that the simulated valuation functions, defined by the algorithm, are lower bounds to the real valuation functions.

By Hall's Theorem (Hall 1935), it is easy to see that there exists a matching of size $\min\{|R|, \frac{1}{4}n^{1/3}\}$ where each agent in R is matched to an item she ranks at the first $\frac{1}{4}n^{1/3}$ positions. Moreover, Y maximizes the social welfare according to the simulated valuation functions. Hence,

$$\text{SW}(Y|\mathbf{v}) \ge \text{SW}(Y|\tilde{\mathbf{v}}) \ge \frac{1}{3}n^{-1/3}\min\Big\{|R|, \frac{1}{4}n^{1/3}\Big\}\,.$$

If $|R| < \frac{1}{4}n^{1/3}$, then $\text{SW}(Y|\mathbf{v}) \ge \frac{1}{3}|R|n^{-1/3}$, and thus

$$\sum_{i\in R} v_i(x_i) \le 3\cdot \text{SW}(Y|\mathbf{v})\,.$$

Otherwise, $\text{SW}(Y|\mathbf{v}) \ge 1/12$, and since $|R| \le n$, we obtain

$$\sum_{i\in R} v_i(x_i) \le 12n^{2/3}\cdot \text{SW}(Y|\mathbf{v})\,.$$

For the second term, we further partition $N\setminus R$ into two sets depending on the position of the x_is. In particular, H is the set of agents $i \in N\setminus R$ who rank x_i at some position $\ell \le n^{1/3}$, and L is the set of remaining agents $i \in (N\setminus R)\setminus H$ (who rank x_i at some position $\ell > n^{1/3}$). Hence,

$$\sum_{i\in N\setminus R} v_i(x_i) = \sum_{i\in H} v_i(x_i) + \sum_{i\in L} v_i(x_i)\,.$$

First consider the agents in H. Since $\max_{i\in N} v_i^* < n^{-1/3}$,

$$\sum_{i\in H} v_i(x_i) \le \max_{i\in N} v_i^*|H| < n^{-1/3}|H|\,.$$

Consider any agent $i \in H$ and any item j that i ranks at some position $\ell \le n^{1/3}$. Since u_i is the value of i for the item she ranks at position $n^{1/3}+1$, we clearly have that $v_i(j) \ge u_i = \tilde{v}_i(j) \ge \frac{1}{2}n^{-1}$. Note that there exists a partial matching of size $|H|$ according to which *all* agents of H are matched to items they rank at the first $n^{1/3}$ positions; e.g., the restriction of X on H. Since Y maximizes the social welfare for the simulated valuation functions, we get

$$\text{SW}(Y|\mathbf{v}) \ge \text{SW}(Y|\tilde{\mathbf{v}}) \ge \frac{1}{2}n^{-1}|H|\,,$$

which immediately implies that

$$\sum_{i\in H} v_i(x_i) \le 2n^{2/3}\cdot \text{SW}(Y|\mathbf{v})\,.$$

Finally, consider the agents in L, and distinguish the following two cases depending on the size of L.

- $|L| \le n^{1/3}$. Since there are at least $n^{1/3}$ different items within the first $n^{1/3}$ positions of each agent in L, by Hall's Theorem, there exists a matching Y' according to which *all* agents in L receive such an item, i.e., every $i \in L$ has (simulated) value at least u_i for the item she gets in Y'. Combining this with the optimality of Y for the simulated valuation functions and the fact that the latter lower bound the real valuation functions, we have
$$\begin{aligned} \text{SW}(Y|\mathbf{v}) &\ge \text{SW}(Y|\tilde{\mathbf{v}}) \ge \text{SW}(Y'|\tilde{\mathbf{v}}) \\ &\ge \sum_{i\in L} u_i \ge \sum_{i\in L} v_i(x_i)\,, \end{aligned}$$
where the last inequality follows by the definition of L.
- $|L| > n^{1/3}$. Denote by S_L the $|S_L| = n^{1/3}$ agents with the highest values u_i among all the agents in L. We may repeat the above argument for S_L instead of L to get $\text{SW}(Y|\mathbf{v}) \ge \sum_{i\in S_L} u_i$. Then,
$$\text{SW}(Y|\mathbf{v}) \ge n^{1/3}\min_{i\in S_L} u_i \ge n^{1/3}\max_{i\in L\setminus S_L} u_i\,.$$
On the other hand, we have
$$\begin{aligned} \sum_{i\in L} v_i(x_i) &\le \sum_{i\in S_L} u_i + (|L|-|S_L|)\max_{i\in L\setminus S_L} u_i \\ &\le \sum_{i\in S_L} u_i + n\max_{i\in L\setminus S_L} u_i \\ &\le (1+n^{2/3})\cdot \text{SW}(Y|\mathbf{v}). \end{aligned}$$

Therefore, the distortion of the algorithm is at most $16n^{2/3}+1$ in case 2. Together with case 1, we obtain the desired bound of $O(n^{2/3}\sqrt{\log n})$. □

6 A General Framework for λ-TSF

Here we generalize λ-TSF, from Section 3, to work for a much broader class of problems, where we are given the ordinal preferences of the agents and access via queries to their cardinal values. We begin with the following general *full information* problem of maximizing an additive objective over a family of combinatorial structures defined on a graph.

Max-on-Graphs: Given a (directed or undirected) weighted graph $G = (U, E, w)$ and a concise description of the set $\mathcal{F} \subseteq 2^E$ of feasible solutions, find a solution $S \in \arg\max_{T\in\mathcal{F}} \sum_{e\in T} w(e)$.

Note that one-sided matching is a special case; G is the complete bipartite graph on N and A, the weight of an edge $\{i, j\}$ is $v_i(j)$, and $\mathcal{F}$ contains the perfect matchings of G.

However, what we are really interested in is the social choice analog of Max-on-Graphs where the weights (defined in terms of the agents' valuation functions) are not given! Instead, we know the ordinal preferences of each agent/node for other nodes (corresponding to items or other agents).

Ordinal-Max-on-Graphs: Here $U = N \cup A$, where N is the set of agents and A is the (possibly empty) set of items; when $A \neq \emptyset$, we assume that G is a bipartite graph with independent sets N, A. Although $G = (U, E)$ is given without the weights, it is assumed that for every $i \in N$ there exists a (private) valuation function $v_i : U \to \mathbb{R}_{\geq 0}$, so that

$$w(e) = \begin{cases} v_i(j), \text{ if } i \in N, j \in A \text{ and } e = \{i, j\} \\ v_i(j) + v_j(i), \text{ if } i, j \in N \text{ and } e = \{i, j\} \\ v_i(j), \text{ if } i, j \in N \text{ and } e = (i, j)\,. \end{cases} \tag{1}$$

We are also given the *ordinal profile* $\succ_{\mathbf{v}} = (\succ_i)_{i\in N}$ induced by $\mathbf{v} = (v_i)_{i\in N}$ and a concise description of the set $\mathcal{F} \subseteq 2^E$ of feasible solutions. The goal is again to find $S \in \arg\max_{T\in\mathcal{F}} \sum_{e\in T} w(e)$.

Notice that for Ordinal-Max-on-Graphs to make sense, $\mathcal{F}$ should be independent of w. E.g., if only sets of weight exactly B are feasible, then it is impossible to find even one feasible set without the exact cardinal information in our disposal. Still, it is clear that the above algorithmic problem is very general and captures a huge number of maximization problems on graphs. Of course, not all such problems have a natural interpretation where the vertices are agents with preferences. Before we state the main result of this section, we give three examples that have been studied in the computational social choice literature.

General Graph Matching: Given an undirected graph $G = (U, E, w)$, find a matching of maximum weight, i.e., $\mathcal{F}$ contains the matchings of G. In its social choice analog, $U = N$ and $w(\cdot)$ is defined according to the second branch of (1). A special case of this problem, when G is a bipartite graph, is the celebrated *two-sided matching* problem (Gale and Shapley 1962; Roth and Sotomayor 1992).

General Resource Allocation: Given a bipartite graph $G = (U_1 \cup U_2, E, w)$, assign each node of U_2 to (only) one neighboring node in U_1 so that the total value of the corresponding edges is maximized. There may be additional combinatorial constraints, e.g., no more than β_i nodes of U_2 may be assigned to node $i \in U_1$. This problem generalizes one-sided matching. In its social choice analog, $U_1 = N$, $U_2 = A$ and $w(\cdot)$ is defined according to the first branch of (1).

Clearing Kidney ℓ-Exchanges: Given a directed graph $G = (U, E, w)$, find a collection of disjoint cycles of length at most ℓ of maximum total weight; see (Abraham, Blum, and Sandholm 2007). In its social choice analog, $U = N$ and $w(\cdot)$ is defined according to the third branch of (1).

We use a variant of λ-TSF, $(\lambda, \mathcal{A})$-TSF, that takes as an additional input an approximation algorithm $\mathcal{A}$ for the problem at hand. There are two main differences from λ-TSF. The simpler one is about the last step; instead of computing a maximum matching, $\mathcal{A}$ is used to compute an (approximately) optimal solution with respect to the simulated valuation functions. The other difference is more subtle. Now we do not want to ask each agent i for her top element of U, but rather for her top element that induces an edge included in some feasible solution. It is not always trivial to find this element for each agent (e.g., think of the variant of general graph matching where we only care about *perfect* matchings), but often it can be done in polynomial time.

For the following theorem, we assume that the optimization problem Π is a special case of Max-on-Graphs with $\max_{T\in\mathcal{F}} |T| = r$. The parameter r allows for a more refined statement. We further assume that we can efficiently check whether an edge e belongs to a feasible solution; if not, there is no guarantee about the running time.

Theorem 7. *Suppose* Π *is as described above. If* $\mathcal{A}$ *is a (polynomial-time)* ρ*-approximation algorithm for* Π *in the full information setting, then* $(\lambda, \mathcal{A})$*-TSF asks* $1+\lambda+\lambda\log r$ *queries and achieves distortion at most* $3\rho\, r^{\frac{1}{\lambda+1}}$ *for the social choice analog of* Π *(in polynomial time).*

For the above problems, we can get the following.

Corollary 3. *By choosing* $\mathcal{A}$ *appropriately,* $(\lambda, \mathcal{A})$*-TSF asks* $1 + \lambda + \lambda \log |U|$ *queries and achieves distortion at most*

- $3\left(\frac{|U|}{2}\right)^{\frac{1}{\lambda+1}}$ *for general graph matching in polynomial time.*
- $3|U_2|^{\frac{1}{\lambda+1}}$ *for general resource allocation.*
- $(4.5+\varepsilon)|U|^{\frac{1}{\lambda+1}}$ *for clearing kidney* 3*-exchanges in polynomial time.*

7 Conclusion and Open Problems

Our work is the first to study the interplay between elicited information and distortion in one-sided matching, as well as other graph problems. We showed several tradeoffs, both in terms of possible distortion guarantees and of inapproximability bounds. Our results suggest that even a small number of queries per agent leads to significant improvements.

A natural future direction would be to try to design algorithms that match the lower bounds of Theorem 4. While we managed to do this for k-well-structured instances, general instances seem to require highly adaptive approaches. Perhaps a first step would be to design an algorithm that outperforms our two-queries algorithm from Section 5 in terms of distortion. Finally, an interesting unexplored avenue is to consider randomized algorithms, either in the selection of the matching or the process of querying the agents.

Acknowledgments

G. Amanatidis has been partially supported by the NWO Veni project VI.Veni.192.153. G. Birmpas has been supported by the ERC Advanced Grant 788893 AMDROMA "Algorithmic and Mechanism Design Research in Online Markets", and the MIUR PRIN project ALGADIMAR "Algorithms, Games, and Digital Markets".

References

Abdulkadiroglu, A.; and Sönmez, T. 2013. Matching markets: Theory and practice. *Advances in Economics and Econometrics* 1: 3–47.

Abraham, D. J.; Blum, A.; and Sandholm, T. 2007. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the 8th ACM conference on Electronic commerce (EC)*, 295–304.

Abramowitz, B.; and Anshelevich, E. 2018. Utilitarians without utilities: Maximizing social welfare for graph problems using only ordinal preferences. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 894–901.

Abramowitz, B.; Anshelevich, E.; and Zhu, W. 2019. Awareness of voter passion greatly improves the distortion of metric social choice. In *Proceedings of the The 15th Conference on Web and Internet Economics (WINE)*, 3–16.

Amanatidis, G.; Birmpas, G.; Filos-Ratsikas, A.; and Voudouris, A. A. 2020a. A Few Queries Go a Long Way: Information-Distortion Tradeoffs in Matching. *CoRR* abs/2009.06543. URL https://arxiv.org/abs/2009.06543.

Amanatidis, G.; Birmpas, G.; Filos-Ratsikas, A.; and Voudouris, A. A. 2020b. Peeking behind the ordinal curtain: Improving distortion via cardinal queries. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 1782–1789.

Anshelevich, E.; Bhardwaj, O.; Elkind, E.; Postl, J.; and Skowron, P. 2018. Approximating optimal social choice under metric preferences. *Artificial Intelligence* 264: 27–51.

Anshelevich, E.; Das, S.; and Naamad, Y. 2013. Anarchy, stability, and utopia: creating better matchings. *Autonomous Agents and Multi-Agent Systems* 26(1): 120–140.

Anshelevich, E.; and Postl, J. 2017. Randomized social choice functions under metric preferences. *Journal of Artificial Intelligence Research* 58: 797–827.

Anshelevich, E.; and Sekar, S. 2016a. Blind, greedy, and random: Algorithms for matching and clustering using only ordinal information. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 390–396.

Anshelevich, E.; and Sekar, S. 2016b. Truthful mechanisms for matching and clustering in an ordinal world. In *International Conference on Web and Internet Economics (WINE)*, 265–278.

Anshelevich, E.; and Zhu, W. 2017. Tradeoffs between information and ordinal approximation for bipartite matching. In *Proceedings of the 10th International Symposium on Algorithmic Game Theory (SAGT)*, 267–279.

Benade, G.; Nath, S.; Procaccia, A. D.; and Shah, N. 2017. Preference elicitation for participatory budgeting. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 376–382.

Bhaskar, U.; Dani, V.; and Ghosh, A. 2018. Truthful and near-optimal mechanisms for welfare maximization in multi-winner elections. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 925–932.

Bogomolnaia, A.; and Moulin, H. 2001. A new solution to the random assignment problem. *Journal of Economic Theory* 100: 295–328.

Boutilier, C.; Caragiannis, I.; Haber, S.; Lu, T.; Procaccia, A. D.; and Sheffet, O. 2015. Optimal social choice functions: A utilitarian view. *Artificial Intelligence* 227: 190–213.

Caragiannis, I.; Nath, S.; Procaccia, A. D.; and Shah, N. 2017. Subset selection via implicit utilitarian voting. *Journal of Artificial Intelligence Research* 58: 123–152.

Filos-Ratsikas, A.; Frederiksen, S. K. S.; and Zhang, J. 2014. Social welfare in one-sided matchings: Random priority and beyond. In *Proceedings of the 7th Symposium of Algorithmic Game Theory (SAGT)*, 1–12.

Gale, D.; and Shapley, L. S. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69(1): 9–15.

Hall, P. 1935. On representatives of subsets. *Journal of the London Mathematical Society* 1(1): 26–30.

Hylland, A.; and Zeckhauser, R. 1979. The efficient allocation of individuals to positions. *Journal of Political economy* 87(2): 293–314.

Klaus, B.; Manlove, D. F.; and Rossi, F. 2016. Matching under preferences. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds., *Handbook of Computational Social Choice*, 333–355. Cambridge University Press.

Kuhn, H. W. 1956. Variants of the Hungarian method for assignment problems. *Naval Research Logistics Quarterly* 3(4): 253–258.

Mandal, D.; Procaccia, A. D.; Shah, N.; and Woodruff, D. P. 2019. Efficient and thrifty voting by any means necessary. In *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 7178–7189.

Mandal, D.; Shah, N.; and Woodruff, D. P. 2020. Optimal communication-distortion tradeoff in voting. In *Proceedings of the 21st ACM Conference on Economics and Computation (EC)*, 795–813.

Procaccia, A. D.; and Rosenschein, J. S. 2006. The distortion of cardinal preferences in voting. In *International Workshop on Cooperative Information Agents (CIA)*, 317–331.

Roth, A. E.; and Sotomayor, M. 1992. Two-sided matching. volume 1 of *Handbook of Game Theory with Economic Applications*, chapter 16, 485 – 541. Elsevier.

Sönmez, T.; and Ünver, M. U. 2011. Matching, allocation, and exchange of discrete resources. In *Handbook of social Economics*, volume 1, 781–852.

Svensson, L.-G. 1999. Strategy-proof allocation of indivisible goods. *Social Choice and Welfare* 16(4): 557–567.

Von Neumann, J.; and Morgenstern, O. 1947. *Theory of games and economic behavior*. Princeton University Press.