Cechlárová, K., Eirinakis, P., Fleiner, T., Magos, D., Manlove, D., Mourtos, I., Ocel'áková, E., and Rastegari, B. (2015) Pareto optimal matchings in many-to-many markets with ties. *Lecture Notes in Computer Science*, 9347, pp. 27-39. (doi:10.1007/978-3-662-48433-3_3)

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/107953/

Deposited on: 20 July 2015

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

Pareto Optimal Matchings in Many-to-Many Markets with Ties *

Katarína Cechlárová[1], Pavlos Eirinakis[2], Tamás Fleiner[3], Dimitrios Magos[4], David Manlove[5], Ioannis Mourtos[2], Eva Oceľáková[1], and Baharak Rastegari[5]

[1] Institute of Mathematics, Faculty of Science, P.J. Šafárik University Košice, Slovakia
[2] Department of Management Science and Technology, Athens University of Economics and Business, Athens, Greece
[3] Budapest University of Technology and Economics Magyar tudósok körútja and MTA-ELTE Egerváry Research Group, Budapest, Hungary
[4] Department of Informatics, Technological Educational Institute of Athens Egaleo, Greece
[5] School of Computing Science, University of Glasgow, Glasgow, UK

Abstract. We consider Pareto-optimal matchings (POMs) in a many-to-many market of applicants and courses where applicants have preferences, which may include ties, over individual courses and lexicographic preferences over sets of courses. Since this is the most general setting examined so far in the literature, our work unifies and generalizes several known results. Specifically, we characterize POMs and introduce the *Generalized Serial Dictatorship Mechanism with Ties (GSDT)* that effectively handles ties via properties of network flows. We show that GSDT can generate all POMs using different priority orderings over the applicants, but it satisfies truthfulness only for certain such orderings. This shortcoming is not specific to our mechanism; we show that any mechanism generating all POMs in our setting is prone to strategic manipulation. This is in contrast to the one-to-one case (with or without ties), for which truthful mechanisms generating all POMs do exist.

Keywords: Pareto optimality, many-to-many matching, serial dictatorship, truthfulness

1 Introduction

We study a many-to-many matching market that involves two finite disjoint sets, a set of applicants and a set of courses. Each applicant finds a subset of

* This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds under Thales grant MIS 380232 (Eirinakis, Magos, Mourtos), by grant EP/K01000X/1 from the Engineering and Physical Sciences Research Council (Manlove, Rastegari), grants VEGA 1/0344/14, 1/0142/15 from the Slovak Scientific grant agency VEGA (Cechlárová), student grant VVGS-PF-2014-463 (Oceľáková) and OTKA grant K108383 (Fleiner). The authors gratefully acknowledge the support of COST Action IC1205 Computational Social Choice.

courses acceptable and has a preference ordering, not necessarily strict, over these courses. Courses do not have preferences. Moreover, each applicant has a quota on the number of courses she can attend, while each course has a quota on the number of applicants it can admit.

A matching is a set of applicant-course pairs such that each applicant is paired only with acceptable courses and the quotas associated with the applicants and the courses are respected. The problem of finding an "optimal" matching given the above market is called the *Course Allocation problem* (CA). Although various optimality criteria exist, *Pareto optimality* (or *Pareto efficiency*) remains the most popular one (see, e.g., [1, 2, 9, 17]). Pareto optimality is a fundamental concept that economists regard as a minimal requirement for a "reasonable" outcome of a mechanism. A matching is a Pareto optimal matching (POM) if there is no other matching in which no applicant is worse off and at least one applicant is better off. Our work examines Pareto optimal many-to-many matchings in the setting where applicants' preferences may include ties.

In the special case where each applicant and course has quota equal to one, our setting reduces to the extensively studied *House Allocation problem* (HA) [13, 1], also known as the *Assignment problem* [11, 5]. Computational aspects of HA have been examined thoroughly [2, 16] and particularly for the case where applicants' preferences are strict. In [2] the authors provide a characterization of POMs in the case of strict preferences and utilize it in order to construct polynomial-time algorithms for checking whether a given matching is a POM and for finding a POM of maximum size. They also show that any POM in an instance of HA with strict preferences can be obtained through the well-known *Serial Dictatorship Mechanism (SDM)* [1]. SDM is a straightforward greedy algorithm that allocates houses sequentially according to some exogenous priority ordering of the applicants, giving each applicant her most-preferred vacant house.

Recently, the above results have been extended in two different directions. The first one [15] considers HA in settings where preferences may include ties. Prior to [15], few works in the literature had considered extensions of SDM to such settings. The difficulty regarding ties, observed already in [18], is that the assignments made in the individual steps of the SDM are not unique, and an unsuitable choice may result in an assignment that violates Pareto optimality. In [6] and [18] an implicit extension of SDM is provided (in the former case for *dichotomous preferences*, where an applicant's preference list comprises a single tie containing all acceptable houses), but without an explicit description of an algorithmic procedure. [15] describe a mechanism called the *Serial Dictatorship Mechanism with Ties (SDMT)* that combines SDM with the use of augmenting paths to ensure Pareto optimality. In an augmentation step, applicants already assigned a house may exchange it for another, equally preferred one, to enable another applicant to take a house that is most preferred given the assignments made so far. They also show that any POM in an instance of HA with ties can be obtained by an execution of SDMT and also describe the so-called *Random Serial Dictatorship Mechanism with Ties (RSDMT)* whose (expected) approximation ratio is $\frac{e}{e-1}$ with respect to the maximum-size POM.

The second direction [9] extends the results of [2] to the many-to-many setting (i.e., CA) with strict preferences, while also allowing for a structure of applicant-wise acceptable sets that is more general than the one implied by quotas; namely, [9] assumes that each applicant selects from a family of course subsets that is downward closed. This work provides a characterization of POMs assuming that the preferences of applicants over sets of courses are obtained from their (strict) preferences over individual courses in a lexicographic manner; using this characterization, it is shown that deciding whether a given matching is a POM can be accomplished in polynomial time. In addition, [9] generalizes SDM to provide the *Generalized Serial Dictatorship (GSD)* mechanism, which can be used to obtain any POM for CA under strict preferences. The main idea of GSD is to allow each applicant to choose not her most preferred set of courses at once but, instead, only one course at a time (i.e., the most preferred among non-full courses that can be added to the courses already chosen). This result is important as the version of SDM where an applicant chooses immediately her most preferred set of course cannot obtain all POMs.

Our contribution. In the current work, we combine the directions appearing in [15] and [9] to explore the many-to-many setting in which applicants have preferences, which may include ties, over individual courses. We extend these preferences to sets of courses lexicographically, since lexicographic set preferences naturally describe human behavior [12], they have already been considered in models of exchange of indivisible goods [9, 10] and also possess theoretically interesting properties including responsiveness [14].

We provide a characterization of POMs in this setting and introduce the *Generalized Serial Dictatorship Mechanism with Ties (GSDT)* that generalizes both SDMT and GSD. SDM assumes a priority ordering over the applicants, according to which applicants are served one by one by the mechanism. Since in our setting applicants can be assigned more than one course, each applicant can return to the ordering several times (up to her quota), each time choosing just one course. The idea of using augmenting paths [15] has to be employed carefully to ensure that during course shuffling no applicant replaces a previously assigned course for a less preferred one. To achieve this, we utilize methods and properties of network flows. Although we prove that GSDT can generate all POMs using different priority orderings over applicants, we also observe that some of the priority orderings guarantee truthfulness whereas some others do not. That is, there may exist priority orderings for which some applicant benefits from misrepresenting her preferences. This is in contrast to SDM and SDMT in the one-to-one case in the sense that all executions of these mechanisms induce truthfulness. This shortcoming however is not specific to our mechanism, since we establish that any mechanism generating all POMs is prone to strategic manipulation by one or more applicants.

Remark. [4] presented a general mechanism for computing Pareto optimal outcomes in hedonic games which includes the many-to-many matching problem with ties. However their mechanism was not presented in a form that is specific to our setting and no explicit bound for the time complexity was given in [4].

applicant	quota	preference list	course	quota
a_1	2	$(\{c_1, c_2\}, \{c_3\}, \emptyset)$	c_1	2
a_2	3	$(\{c_2\}, \{c_1, c_3\}, \emptyset)$	c_2	1
a_3	2	$(\{c_3\}, \{c_2\}, \{c_1\})$	c_3	1

Table 1. An instance I of CA.

Organization of the paper. In Section 2 we define our notation and terminology. The characterization is provided in Section 3, while GSDT is presented in Section 4. A discussion on applicants' incentives in GSDT is provided in Section 5. Missing proofs can be found in the full version of this paper [8].

2 Preliminary definitions of notation and terminology

Let $A = \{a_1, a_2, \cdots, a_{n_1}\}$ be the set of applicants, $C = \{c_1, c_2, \cdots, c_{n_2}\}$ the set of courses and $[i]$ denote the set $\{1, 2, \ldots, i\}$. Each applicant a has a quota $b(a)$ that denotes the maximum number of courses a can accommodate into her schedule, and likewise each course c has a quota $q(c)$ that denotes the maximum number of applicants it can admit. Each applicant finds a subset of courses acceptable and has a transitive and complete preference ordering, not necessarily strict, over these courses. We write $c \succ_a c'$ to denote that applicant a *(strictly) prefers* course c to course c', and $c \simeq_a c'$ to denote that a is *indifferent between* c and c'. We write $c \succeq_a c'$ to denote that a either prefers c to c' or is indifferent between them, and say that a *weakly prefers* c to c'.

Because of indifference, each applicant divides her acceptable courses into *indifference classes* such that she is indifferent between the courses in the same class and has a strict preference over courses in different classes. Let C_t^a denote the t'th indifference class, or *tie*, of applicant a where $t \in [n_2]$. We assume that $C_t^a = \emptyset$ implies $C_{t'}^a = \emptyset$ for all $t' > t$. Let the preference list of any applicant a be the tuple of sets C_t^a, i.e., $P(a) = (C_1^a, C_2^a, \cdots, C_{n_2}^a)$; occasionally we consider $P(a)$ to be a set itself and write $c \in P(a)$ instead of $c \in C_t^a$ for some t. We denote by $\mathcal{P}$ the joint preference profile of all applicants, and by $\mathcal{P}(-a)$ the joint profile of all applicants except a. Under these definitions, an instance of CA is denoted by $I = (A, C, \mathcal{P}, b, q)$. Such an instance appears in Table 1.

A *(many-to-many) assignment* μ is a subset of $A \times C$. For $a \in A$, $\mu(a) = \{c \in C : (a, c) \in \mu\}$ and for $c \in C$, $\mu(c) = \{a \in A : (a, c) \in \mu\}$. An assignment μ is a *matching* if $\mu(a) \subseteq P(a)$, $|\mu(a)| \leq b(a)$ for each $a \in A$ and $|\mu(c)| \leq q(c)$ for each $c \in C$. We say that a is *exposed* if $|\mu(a)| < b(a)$, and is *full* otherwise. Analogous definitions of exposed and full hold for courses.

For an applicant a and a set of courses S, we define the *generalized characteristic vector* $\chi_a(S)$ as the vector $(|S \cap C_1^a|, |S \cap C_2^a|, \ldots, |S \cap C_{n_2}^a|)$. We assume that for any two sets of courses S and U, a prefers S to U if and only if $\chi_a(S) >_{lex} \chi_a(U)$, i.e., if and only if there is an indifference class C_t^a such that $|S \cap C_t^a| > |U \cap C_t^a|$ and $|S \cap C_{t'}^a| = |U \cap C_{t'}^a|$ for all $t' < t$. If a neither prefers S to U nor U to S, then she is indifferent between S and U. We write $S \succ_a U$ if a prefers S to U, $S \simeq_a U$ if a is indifferent between S and U, and $S \succeq_a U$ if a weakly prefers S to U.

A matching μ is a *Pareto optimal matching (POM)* if there is no other matching in which some applicant is better off and none is worse off. Formally, μ is Pareto optimal if there is no matching μ' such that $\mu'(a) \succeq_a \mu(a)$ for all $a \in A$, and $\mu'(a') \succ_{a'} \mu(a')$ for some $a' \in A$. If such a μ' exists, we say that μ' *Pareto dominates* μ.

A *deterministic mechanism* ϕ maps an instance to a matching, i.e. $\phi : I \mapsto \mu$ where I is a CA instance and μ is a matching in I. A *randomized mechanism* ϕ maps an instance to a distribution over possible matchings. Applicants' preferences are private knowledge and an applicant may prefer not to reveal her preferences truthfully. A deterministic mechanism is *truthful* if all applicants always finds it best to declare their true preferences, no matter what other applicants declare. A randomized mechanism ϕ is *universally truthful* if it is a probability distribution over deterministic truthful mechanisms.

3 Characterizing Pareto optimal matchings

Manlove [16, Sec. 6.2.2.1] provided a characterization of Pareto optimal matchings in HA with preferences that may include indifference. He defined three different types of *coalitions* with respect to a given matching such that the existence of either means that a subset of applicants can trade among themselves (possibly using some exposed course) and ensure that, at the end, no one is worse off and at least one applicant is better off. He also showed that if no such coalition exists, then the matching is guaranteed to be Pareto optimal. We show that this characterization extends to the many-to-many setting, although the proof is more complex and involved than in the one-to-one setting.

In what follows we assume that in each sequence $\mathfrak{C}$ no applicant or course appears more than once.

An *alternating path coalition* w.r.t. μ comprises a sequence $\mathfrak{C} = \langle c_{j_0}, a_{i_0}, c_{j_1}, a_{i_1}, \ldots, c_{j_{r-1}} a_{i_{r-1}}, c_{j_r} \rangle$ where $r \geq 1$, $c_{j_k} \in \mu(a_{i_k})$ $(0 \leq k \leq r-1)$, $c_{j_k} \notin \mu(a_{i_{k-1}})$ $(1 \leq k \leq r)$, a_{i_0} is full, and c_{j_r} is an exposed course. Furthermore, a_{i_0} prefers c_{j_1} to c_{j_0} and, if $r \geq 2$, a_{i_k} weakly prefers $c_{j_{k+1}}$ to c_{j_k} $(1 \leq k \leq r-1)$.

An *augmenting path coalition* w.r.t. μ comprises a sequence $\mathfrak{C} = \langle a_{i_0}, c_{j_1}, a_{i_1}, \ldots, c_{j_{r-1}} a_{i_{r-1}}, c_{j_r} \rangle$ where $r \geq 1$, $c_{j_k} \in \mu(a_{i_k})$ $(1 \leq k \leq r-1)$, $c_{j_k} \notin \mu(a_{i_{k-1}})$ $(1 \leq k \leq r)$, a_{i_0} is an exposed applicant, and c_{j_r} is an exposed course. Furthermore, a_{i_0} finds c_{j_1} acceptable and, if $r \geq 2$, a_{i_k} weakly prefers $c_{j_{k+1}}$ to c_{j_k} $(1 \leq k \leq r-1)$.

A *cyclic coalition* w.r.t. μ comprises a sequence $\mathfrak{C} = \langle c_{j_0}, a_{i_0}, c_{j_1}, a_{i_1}, \ldots, c_{j_{r-1}}, a_{i_{r-1}} \rangle$ where $r \geq 2$, $c_{j_k} \in \mu(a_{i_k})$ $(0 \leq k \leq r-1)$, and $c_{j_k} \notin \mu(a_{i_{k-1}})$ $(1 \leq k \leq r)$. Furthermore, a_{i_0} prefers c_{j_1} to c_{j_0} and a_{i_k} weakly prefers $c_{j_{k+1}}$ to c_{j_k} $(1 \leq k \leq r-1)$. (All subscripts are taken modulo r when reasoning about cyclic coalitions).

We define an *improving coalition* to be an alternating path coalition, an augmenting path coalition or a cyclic coalition. Given an improving coalition $\mathfrak{C}$, the matching

$$\mu^{\mathfrak{C}} = (\mu \setminus \{(a_{i_k}, c_{j_k}) : \delta \leq k \leq r-1\}) \cup \{(a_{i_k}, c_{j_{k+1}}) : 0 \leq k \leq r-1\} \quad (1)$$

is defined to be the matching obtained from μ by *satisfying* $\mathfrak{C}$ ($\delta = 1$ in the case that $\mathfrak{C}$ is an augmenting path coalition, otherwise $\delta = 0$).

The following theorem gives a necessary and sufficient condition for a matching to be Pareto optimal.

Theorem 1. *Given a CA instance I, a matching μ is a Pareto optimal matching in I if and only if μ admits no improving coalition.*

4 Constructing Pareto optimal matchings

We propose an algorithm for finding a POM in an instance of CA, which is in a certain sense a generalization of Serial Dictatorship thus named 'Generalized Serial Dictatorship with ties' (GSDT). The algorithm starts by setting the quotas of all applicants to 0 and those of courses are set at the original values given by q. At each stage i, the algorithm selects a single applicant whose original capacity has not been reached, and increases only her capacity by 1. The algorithm terminates after $B = \sum_{a \in A} b(a)$ stages, i.e., once the original capacities of all applicants have been reached. In that respect, the algorithm assumes a 'multisequence' $\Sigma = (a^1, a^2, \ldots, a^B)$ of applicants such that each applicant a appears $b(a)$ times in Σ; e.g., for the instance of Table 1 and the sequence $\Sigma = (a_1, a_1, a_2, a_2, a_3, a_2, a_3)$, the vector of capacities evolves as follows:

$$(0,0,0),(1,0,0),(2,0,0),(2,1,0),(2,2,0),(2,2,1),(2,3,1),(2,3,2).$$

Let us denote the vector of applicants' capacities in stage i by b^i, i.e., b^0 is the all-zeroes vector and $b^B = b$. Clearly, each stage corresponds to an instance I^i similar to the original instance except for the capacities vector b^i. At each stage i, our algorithm obtains a matching μ^i for the instance I^i. The single matching of stage 0—the empty matching, is a POM in I^0. The core idea is to modify μ^{i-1} in such way that if μ^{i-1} is a POM with respect to I^{i-1} then μ^i is a POM with respect to I^i. To achieve this, the algorithm relies on the following flow network.

Consider the digraph $D = (V, E)$. Its node set is $V = A \cup T \cup C \cup \{\sigma, \tau\}$ where σ and τ are the source and the sink and vertices in T correspond to the ties in the preference lists of all applicants; i.e., T has a node (a, t) per applicant a and tie t such that $C_t^a \neq \emptyset$. Its arc set is $E = E_1 \cup E_2 \cup E_3 \cup E_4$ where $E_1 = \{(\sigma, a) : a \in A\}$, $E_2 = \{(a, (a, t)) : a \in A, C_t^a \neq \emptyset\}$, $E_3 = \{((a, t), c) : c \in C_t^a\}$ and $E_4 = \{(c, \tau) : c \in C\}$. The graph D for the instance of Table 1 appears in Figure 1.

Using digraph $D = (V, E)$, we obtain a flow network N^i at each stage i of the algorithm, i.e., a network corresponding to instance I^i, by appropriately varying the capacities of the arcs. (For an introduction on network flow algorithms see, e.g., [3].) The capacity of each arc in E_3 is always 1 (since each course may be received at most once by each applicant) and the capacity of an arc $e = (c, \tau) \in E_4$ is always $q(c)$. The capacities of all arcs in $E_1 \cup E_2$ are initially 0 and, at stage i, the capacities of only certain arcs associated with applicant a^i are increased by 1. For this reason, for each applicant a we use the variable $curr(a)$ that indicates her 'active' tie; initially, $curr(a)$ is set to 1 for all $a \in A$.

In stage i, the algorithm computes a maximum flow f^i whose saturated arcs in E_3 indicate the corresponding matching μ^i. The algorithm starts with $f^0 = 0$

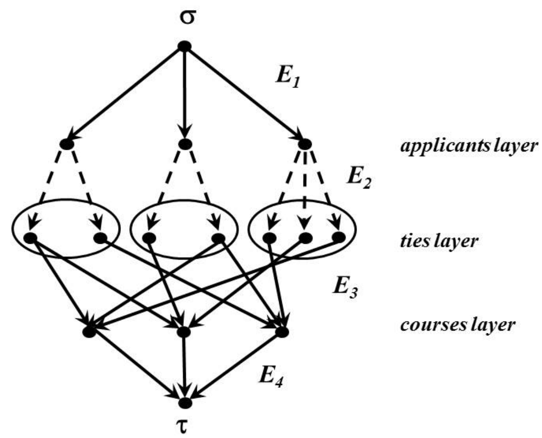

Fig. 1. Digraph D for the instance I from Table 1. An oval encircles all the vertices of T that correspond to the same applicant.

and $\mu^0 = \emptyset$. Let the applicant $a^i \in A$ be a copy of applicant a considered in stage i. The algorithm increases by 1 the capacity of arc $(\sigma, a) \in E_1$ (i.e., the applicant is allowed to receive an additional course). It then examines the tie $curr(a)$ to check whether the additional course can be received from tie $curr(a)$. To do this, the capacity of arc $(a, (a, curr(a))) \in E_2$ is increased by 1. The network in stage i where tie $curr(a^i)$ is examined is denoted by $N^{i,curr(a^i)}$. If there is an augmenting $\sigma - \tau$ path in this network, the algorithm augments the current flow f^{i-1} to obtain f^i, accordingly augments μ^{i-1} to obtain μ^i (i.e., it sets μ^i to the symmetric difference of μ^{i-1} and all pairs (a, c) for which there is an arc $((a, t), c)$ in the augmenting path) and proceeds to the next stage. Otherwise, it decreases the capacity of $(a, (a, curr(a)))$ by 1 (but not the capacity of arc (σ, a)) and it increases $curr(a)$ by 1 to examine the next tie of a; if all (non-empty) ties have been examined, the algorithm proceeds to the next stage without augmenting the flow. Note that an augmenting $\sigma - \tau$ path in the network $N^{i,curr(a^i)}$ corresponds to an augmenting path coalition in μ^{i-1} with respect to I^i

A formal description of GSDT is provided by Algorithm 1, where $w(e)$ denotes the capacity of an arc $e \in E$ and $\oplus$ denotes the operation of augmenting along an augmenting path (either relative to a flow or a matching). Observe that all arcs in E_2 are saturated, except for the arc corresponding to the current applicant and tie, thus any augmenting path has one arc from each of E_1, E_2 and E_4 and all other arcs from E_3; as a consequence, the number of courses each applicant receives at stage i in any tie cannot decrease at any subsequent step. Also, μ^i dominates μ^{i-1} with respect to instance I^i if and only if there is a flow in N^i that saturates all arcs in E_2.

To prove the correctness of GSDT, we need two intermediate lemmas. Let $e_t \in \mathbb{R}^{n_2}$ be the vector having 1 at entry t and 0 elsewhere.

Lemma 1. *Let $N^{i,t}$ be the network at stage i while tie t of applicant a^i is examined. Then, there is an augmenting path with respect to f^{i-1} in $N^{i,t}$ if and*

Algorithm 1: Producing a POM for any instance of CA

Input: an instance I of CA and a multisequence Σ
$f^0 := 0$; $\mu^0 := \emptyset$;
for each $a \in A$, $curr(a) := 1$;
for $i = 1, 2, \ldots, B$ **do**
{
    consider the applicant $a = a^i$;
    $w(\sigma, a)$++;
    $\mathfrak{P} := \emptyset$;
    while $\mathfrak{P} = \emptyset$ **and** $curr(a) \leq n_2$ **and** $C^a_{curr(a)} \neq \emptyset$ **do**
    {
        $w(a, (a, curr(a)))$++;
        $\mathfrak{P} :=$ augmenting path in $N^{i,curr(a)}$ with respect to f^{i-1};
        if $\mathfrak{P} = \emptyset$ **then** { $w(a, (a, curr(a)))$- -; $curr(a)$++};
    }
    if $\mathfrak{P} \neq \emptyset$ **then** { $f^i := f^{i-1} \oplus \mathfrak{P}$;
        $\mu^i := \mu^{i-1} \oplus \{(a, c) : ((a, t), c) \in \mathfrak{P}$ for some tie $t\}$; }
    otherwise { $f^i := f^{i-1}$; $\mu^i := \mu^{i-1}$; }
}
return μ^B;

only if there is a matching μ such that

$$\chi_a(\mu(a)) = \chi_a(\mu^{i-1}(a)) \text{ for each } a \neq a^i \text{ and } \chi_{a^i}(\mu(a^i)) = \chi_{a^i}(\mu^{i-1}(a^i)) + e_t.$$

Lemma 2. *Let $S \succeq_a U$ and $|S| \geq |U|$. If c_S and c_U denote a least preferred course of applicant a in S and U, respectively, then $S \backslash \{c_S\} \succeq_a U \backslash \{c_U\}$.*

Theorem 2. *For each i, the matching μ^i obtained by GSDT is a POM for instance I^i.*

Proof. We apply induction on i. Clearly, $\mu^0 = \emptyset$ is the single matching in I^0 and hence a POM in I^0. We assume that μ^{i-1} is a POM in I^{i-1} and prove that μ^i is a POM in I^i.

Assuming to the contrary that μ^i is not a POM in I^i implies that there is a matching ξ in I^i that dominates μ^i. Then, for all $a \in A$, $\xi(a) \succeq_a \mu^i(a) \succeq_a \mu^{i-1}(a)$. Since the capacities of all applicants in I^i are as in I^{i-1} except for the capacity of a^i that has been increased by 1, for all $a \in A \setminus \{a_i\}$, $|\xi(a)|$ does not exceed the capacity of a in instance I^{i-1}, namely $b^{i-1}(a)$, while $|\xi(a^i)|$ may exceed $b^{i-1}(a^i)$ by at most 1.

Moreover, it holds that $|\xi(a^i)| \geq |\mu^i(a^i)|$. Assuming to the contrary, that $|\xi(a^i)| < |\mu^i(a^i)|$ yields that ξ is feasible also in instance I^{i-1}. In addition, $|\xi(a^i)| < |\mu^i(a^i)|$ implies that it cannot be $\xi(a^i) \simeq_{a^i} \mu^i(a^i)$ thus, together with $\xi(a^i) \succeq_{a^i} \mu^i(a^i) \succeq_{a^i} \mu^{i-1}(a^i)$, it yields $\xi(a^i) \succ_{a^i} \mu^i(a^i) \succeq_{a^i} \mu^{i-1}(a^i)$. But then, ξ dominates μ^{i-1} in I^{i-1}, a contradiction to μ^{i-1} being a POM in I^{i-1}.

Let us first examine the case in which GSDT enters the 'while' loop and finds an augmenting path, hence μ^i dominates μ^{i-1} in I^i only with respect to applicant a^i that receives an additional course. This is one of her worst courses in $\mu^i(a^i)$ denoted as c_μ. Let c_ξ be a worst course for a^i in $\xi(a^i)$. Let also ξ' and μ' denote $\xi \setminus \{(a^i, c_\xi)\}$ and $\mu^i \setminus \{(a^i, c_\mu)\}$, respectively. Observe that both ξ' and μ' are feasible in I^{i-1}, while having shown that $|\xi(a^i)| \geq |\mu^i(a^i)|$ implies through Lemma 2 that ξ' weakly dominates μ' which in turn weakly dominates μ^{i-1} by Lemma 1. Since μ^{i-1} is a POM in I^{i-1}, $\xi'(a) \simeq_a \mu'(a) \simeq_a \mu^{i-1}(a)$ for all $a \in A$, therefore ξ dominates μ^i only with respect to a^i and $c_\xi \succ_{a^i} c_\mu$. Overall, $\xi(a) \simeq_a \mu^i(a) \simeq_a \mu^{i-1}(a)$ for all $a \in A \setminus \{a^i\}$ and $\xi(a^i) \succ_{a^i} \mu^i(a^i) \succ_{a^i} \mu^{i-1}(a^i)$.

Let t_ξ and t_μ be the ties of applicant a^i containing c_ξ and c_μ, respectively, where $t_\xi < t_\mu$ because $c_\xi \succ_{a^i} c_\mu$. Then, Lemma 1 implies that there is a path augmenting f^{i-1} (i.e., the flow corresponding to μ^{i-1}) in the network N^{i,t_ξ}. Let also t' be the value of $curr(a^i)$ at the beginning of stage i. Since we examine the case where GSDT enters the 'while' loop and finds an augmenting path, $C_{t'}^{a^i} \neq \emptyset$. Thus, t' indexes the least preferred tie from which a^i has a course in μ^{i-1}, the same holding for ξ' since $\xi'(a^i) \simeq_{a^i} \mu^{i-1}(a^i)$. Because ξ' is obtained by removing from a^i its worst course in $\xi(a^i)$, that course must belong to a tie of index no smaller that t', i.e., $t' \leq t_\xi$. This together with $t_\xi < t_\mu$ yield $t' \leq t_\xi < t_\mu$, which implies that GSDT should have obtained ξ instead of μ^i at stage i, a contradiction.

It remains to examine the cases where, at stage i, GSDT does not enter the 'while' loop or enters it but finds no augmenting path. For both these cases, $\mu^i = \mu^{i-1}$, thus ξ dominating μ^i means that ξ is not feasible in I^{i-1} (since it would then also dominate μ^{i-1}). Then, it holds that $|\xi(a^i)|$ exceeds $b^{i-1}(a^i)$ by 1, thus $|\xi(a^i)| > |\mu^i(a^i)|$ yielding $\xi(a^i) \succ_{a^i} \mu^i(a^i)$. Let t_ξ be defined as above and t' now be the most preferred tie from which a^i has more courses in ξ than in μ^i. Clearly, $t' \leq t_\xi$ since t_ξ indexes the least preferred tie from which a^i has a course in ξ. If $t' < t_\xi$ then the matching ξ', defined as above, is feasible in I^{i-1} and dominates μ^{i-1} because $\xi'(a^i) \succ_{a^i} \mu^{i-1}(a^i)$, a contradiction; the same holds if $t' = t_\xi$ and a^i has in ξ at least two more courses from t_ξ than in μ^i. Otherwise, $t' = t_\xi$ and a^i has in ξ exactly one more course from t_ξ than in μ^i; that, together with $|\xi(a^i)| > |\mu^i(a^i)|$ and the definition of t_ξ, implies that the index of the least preferred tie from which a^i has a course in μ^{i-1} and, therefore, the value of $curr(a^i)$ in the beginning of stage i, is at most t'. But then GSDT should have obtained ξ instead of μ^i at stage i, a contradiction. □

The following statement is now direct.

Corollary 1. *GSDT produces a POM for instance I.*

To derive the complexity bound for GSDT, let us denote by L the length of the preference profile in I, i.e. the total number of courses in the preference lists of all applicants. Notice that $|E_3| = L$ and neither the size of any matching in I nor the total number of ties in all preference lists exceeds L.

Within one stage, several searches in the network might be needed to find a tie of the active applicant for which the current flow can be augmented. However, one tie is unsuccessfully explored at most once, hence each search either augments the flow thus adding a pair to the current matching or moves to the next tie. So the total number of searches performed by the algorithm is bounded by the size of the obtained matching plus the number of ties in the preference profile, i.e. it is $O(L)$. A search requires a number of steps that remains linear in the number of arcs in the current network (i.e., $N^{i,curr(a^i)}$), but as at most one arc per E_1, E_2 and E_4 is used, any search needs $O(|E_3|) = O(L)$ steps. This leads to a complexity bound $O(L^2)$ for GSDT.

Next we show that GSDT can produce any POM.

Theorem 3. *Given a CA instance I and a POM μ, there exists a suitable priority ordering over applicants Σ given which GSDT can produce μ.*

5 Truthfulness of mechanisms for finding POMs

It is well-known that the SDM for HA is truthful, regardless of the given priority ordering over applicants. We will show shortly that GSDT is not necessarily truthful, but first prove that this property does hold for some priority orderings over applicants.

Theorem 4. *GSDT is truthful given Σ if, for each applicant a, all occurrences of a in Σ are consecutive.*

Proof. W.l.o.g. let the applicants appear in Σ in the following order

$$\underbrace{a_1, a_1, \ldots, a_1}_{b(a_1)\text{-times}}, \underbrace{a_2, a_2, \ldots, a_2}_{b(a_2)\text{-times}} \ldots, \underbrace{a_{i-1}, a_{i-1}, \ldots, a_{i-1}}_{b(a_{i-1})\text{-times}}, \underbrace{a_i, a_i, \ldots, a_i}_{b(a_i)\text{-times}}, \ldots$$

Assume to the contrary that some applicant benefits from misrepresenting her preferences. Let a_i be the first such applicant in Σ who reports $P'(a_i)$ instead of $P(a_i)$ in order to benefit and $\mathcal{P}' = (P'(a_i), \mathcal{P}(-a_i))$. Let μ denote the matching returned by GSDT using ordering Σ on instance $I = (A, C, \mathcal{P}, b, q)$ (i.e. the instance in which applicant a_i reports truthfully) and ξ the matching returned by GSDT using Σ but on instance $I' = (A, C, \mathcal{P}', b, q)$. Let $s = (\Sigma_{\ell<i} b(a_\ell)) + 1$, i.e., s is the first stage in which our mechanism considers applicant a_i. Let j be the first stage of GSDT such that a_i prefers ξ^j to μ^j, where $s \leq j < s + b(a_i)$.

Given that applicants $a_1, \ldots, a_{i-1}$ report the same in I as in I' and all their occurrences in Σ are before stage j, Lemma 1 yields $\mu^j(a_\ell) \simeq_{a_\ell} \xi^j(a_\ell)$ for $\ell = 1, 2, \ldots, i-1$. Also $\mu^j(a_\ell) = \xi^j(a_\ell) = \emptyset$ for $\ell = i+1, i+2, \ldots, n_1$, since no such applicant has been considered before stage j. But then, all applicants apart from a_i are indifferent between μ^j and ξ^j, therefore a_i preferring ξ^j to μ^j implies that μ^j is not a POM in I^j, a contradiction to Theorem 2. $\square$

The next result then follows directly from Theorem 4.

Corollary 2. *GSDT is truthful if all applicants have quota equal to one.*

There are priority orderings for which an applicant may benefit from misreporting her preferences, even if preferences are strict. This phenomenon has also been observed in a slightly different context [7]. Let us also provide an example.

$a_1 : c_1 \succ c_2$	$a_1 : c_1 \succ c_2$	$a_1 : c_2 \succ c_1$	$a_1 : c_2 \succ c_1$
$a_2 : c_1 \succ c_2$	$a_2 : c_1$	$a_2 : c_1$	$a_2 : c_1 \succ c_2$
I_1 with μ_1	I_2 with μ_2	I_3 with μ_2	I_4

Fig. 2. Four instances of CA used in the proof of Theorem 5. In all four instances $b(a_1) = 2$, $b(a_2) = 1$, $q(c_1) = q(c_2) = 1$. For each of instances I_1 to I_3, a matching is indicated using circles in applicants' preference lists.

Example 1. Consider a setting with applicants a_1 and a_2 and courses c_1 and c_2, for which $b(a_1) = 2$, $b(a_2) = 1$, $q(c_1) = 1$, and $q(c_2) = 1$. Let I be an instance in which $c_2 \succ_{a_1} c_1$ and a_2 finds only c_1 acceptable. This setting admits two POMs, namely $\mu_1 = \{(a_1, c_2), (a_2, c_1)\}$ and $\mu_2 = \{(a_1, c_1), (a_1, c_2)\}$. GSDT returns μ_1 for $\Sigma = (a_1, a_2, a_1)$. If a_1 misreports by stating that she prefers c_1 to c_2, GSDT returns μ_2 instead of μ_1. Since $\mu_2 \succ_{a_1} \mu_1$, GSDT is not truthful given Σ.

The above observation seems to be a deficiency of GSDT. We conclude by showing that no mechanism capable of producing all POMs is immune to this shortcoming.

Theorem 5. *There is no universally truthful randomized mechanism that produces all POMs in CA, even if applicants' preferences are strict and all courses have quota equal to one.*

Proof. The instance I_1 in Figure 2 admits three POMs, namely $\mu_1 = \{(a_1, c_1), (a_2, c_2)\}$, $\mu_2 = \{(a_1, c_1), (a_1, c_2)\}$ and $\mu_3 = \{(a_1, c_2), (a_2, c_1)\}$. Assume a randomized mechanism ϕ that produces all these matchings. Therefore, there must be a deterministic realization of it, denoted as ϕ^D, that returns μ_1 given I_1. Let us examine the outcome of ϕ^D under the slightly different applicants' preferences shown in Figure 2, bearing in mind that ϕ^D is truthful.

– Under I_2, ϕ^D must return μ_2. The only other POM under I_2 is μ_3, but if ϕ^D returns μ_3 then a_2 under I_1 has an incentive to lie and declare only c_1 acceptable (as in I_2).
– Under I_3, ϕ^D must return μ_2. The only other POM under I_3 is μ_3, but if ϕ^D returns μ_3 then a_1 under I_3 has an incentive to lie and declare that she prefers c_1 to c_2 (as in I_2).

I_4 admits two POMs, namely μ_2 and μ_3. If ϕ^D returns μ_2, then a_1 under I_1 has an incentive to lie and declare that she prefers c_2 to c_1 (as in I_4). If ϕ^D returns μ_3, then a_2 under I_3 has an incentive to lie and declare c_2 acceptable—in addition to c_1—and less preferred than c_1 (as in I_4). Thus overall ϕ^D cannot return a POM under I_4 while maintaining truthfulness. □

6 Future work

A particularly important problem is to investigate the expected size of the matching produced by the randomized version of GSDT. It is also interesting to characterize priority orderings that induce truthful-telling in GSDT. Should this be possible, it would be interesting to compute the expected size of the matching produced by a randomized GSDT in which the randomization is taken over the priority orderings that give rise to truthfulness.

References

1. A. Abdulkadiroğlu and T. Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–701, 1998.
2. D.J. Abraham, K. Cechlárová, D.F. Manlove, and K. Mehlhorn. Pareto optimality in house allocation problems. In *Proc. ISAAC '04*, volume 3341 of *LNCS*, pages 3–15. Springer, 2004.
3. R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
4. H. Aziz, F. Brandt, and P. Harrenstein. Pareto optimality in coalition formation. *Games and Economic Behavior*, 82:562 – 581, 2013.
5. A. Bogomolnaia and H. Moulin. A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2):295 – 328, 2001.
6. A. Bogomolnaia and H. Moulin. Random matching under dichotomous preferences. *Econometrica*, 72(1):257–279, 2004.
7. E. Budish and E. Cantillon. The multi-unit assignment problem: Theory and evidence from course allocation at harvard. *American Economic Review*, 102(5):2237–71, 2012.
8. K. Cechlárová, P. Eirinakis, T. Fleiner, D. Magos, D. Manlove, I. Mourtos, E. Oceľáková, and B. Rastegari. Pareto optimal matchings in many-to-many markets with ties. Technical Report 1507.02866, 2015. Available from http://arxiv.org/abs/1507.02866.
9. K. Cechlárová, P. Eirinakis, T. Fleiner, D. Magos, I. Mourtos, and E. Potpinková. Pareto optimality in many-to-many matching problems. *Discrete Optimization*, 14(0):160 – 169, 2014.
10. E. Fujita, J. Lesca, A. Sonoda, T. Todo, and M. Yokoo. A complexity approach for core-selecting exchange with multiple indivisible goods under lexicographic preferences. In *Proc. AAAI '15*, 2015.
11. P. Gärdenfors. Assignment problem based on ordinal preferences. *Management Science*, 20(3):331–340, 1973.
12. G. Gigerenzer and D. G. Goldstein. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4):650–669, 1996.
13. A. Hylland and R. Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political Economy*, 87(2):293–314, 1979.
14. B. Klaus and E. Miyagawa. Strategy-proofness, solidarity, and consistency for multiple assignment problems. *Int J Game Theory*, 30:421–435, 2001.
15. P. Krysta, D. Manlove, B. Rastegari, and J. Zhang. Size versus truthfulness in the House Allocation problem. Technical Report 1404.5245, 2014. A shorter version appeared in the Proceedings of EC'14.
16. D.F. Manlove. *Algorithmics of Matching Under Preferences*. World Scientific, 2013.
17. D. Saban and J. Sethuraman. The complexity of computing the random priority allocation matrix. In *Proc. WINE '13*, volume 8289 of *LNCS*, page 421. Springer, 2013.
18. L. G. Svensson. Queue allocation of indivisible goods. *Social Choice and Welfare*, 11(4):323–330, 1994.