



Discrete Optimization

Mathematical models for stable matching problems with ties and incomplete lists

Maxence Delorme^{a,*}, Sergio García^a, Jacek Gondzio^a, Jörg Kalcsics^a, David Manlove^b, William Pettersson^b

^a School of Mathematics, University of Edinburgh, United Kingdom

^b School of Computing Science, University of Glasgow, United Kingdom



ARTICLE INFO

Article history:

Received 7 September 2018

Accepted 13 March 2019

Available online 18 March 2019

Keywords:

Assignment

Stable Marriage problem

Hospitals/Residents problem

Ties and Incomplete lists

Exact algorithms

ABSTRACT

We present new integer linear programming (ILP) models for \mathcal{NP} -hard optimisation problems in instances of the Stable Marriage problem with Ties and Incomplete lists (SMTI) and its many-to-one generalisation, the Hospitals/Residents problem with Ties (HRT). These models can be used to efficiently solve these optimisation problems when applied to (i) instances derived from real-world applications, and (ii) larger instances that are randomly-generated. In the case of SMTI, we consider instances arising from the pairing of children with adoptive families, where preferences are obtained from a quality measure of each possible pairing of child to family. In this case, we seek a maximum weight stable matching. We present new algorithms for preprocessing instances of SMTI with ties on both sides, as well as new ILP models. Algorithms based on existing state-of-the-art models only solve 6 of our 22 real-world instances within an hour per instance, and our new models incorporating dummy variables and constraint merging, together with preprocessing and a warm start, solve all 22 instances within a mean runtime of a minute. For HRT, we consider instances derived from the problem of assigning junior doctors to foundation posts in Scottish hospitals. Here, we seek a maximum size stable matching. We show how to extend our models for SMTI to HRT and reduce the average running time for real-world HRT instances by two orders of magnitude. We also show that our models outperform by a wide margin all known state-of-the-art models on larger randomly-generated instances of SMTI and HRT.

© 2019 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

1.1. Background

In a stable matching problem, we are given a set of agents, each of whom ranks all the others in strict order of preference, indicating their level of desire to be matched to each other. A solution of the problem is a pairing of all agents such that no two agents form a *blocking pair*, i.e., a pair that are not currently matched together, but would prefer to be matched to each other rather than to their currently assigned partners.

Without any other constraints, this problem is known as the Stable Roommates (SR) problem (Gale & Shapley, 1962; Gusfield & Irving, 1989), and the objective is to partition the n agents into $n/2$

pairs (e.g., doubles in a tennis tournament) such that no blocking pair exists.

The Stable Marriage problem (SM) is a bipartite restriction of SR, where the agents are split into equal-sized sets of men and women, and it is assumed that men only find women acceptable and vice versa. This problem was first introduced by Gale and Shapley (1962), who also gave a linear-time algorithm for finding a stable matching.

It is not always desirable, or even possible, to have every agent express a preference over all other agents. In the Stable Marriage problem with Incomplete lists (SMI), agents can identify potential partners as being unacceptable, meaning that they would rather be unmatched than matched to such agents, and a slight modification of the Gale–Shapley algorithm will find a stable matching in linear time (Gusfield and Irving, 1989, Section 1.4.2). It turns out that all stable matchings in a given instance of SMI have the same size (Gale & Sotomayor, 1985).

In many applications it is not realistic to expect that agents have sufficient information to enable them to rank their acceptable

* Corresponding author.

E-mail address: maxence.delorme@ed.ac.uk (M. Delorme).

potential partners in strict order of preference. In reality, preference lists may include ties, where a tie indicates a set of agents that are equally desirable. This gives rise to another variant of SM known as the Stable Marriage problem with Ties (SMT) (Irving, 1994). It is known that resolving indifference by employing tie-breaking is not a good strategy, since it over-constrains the problem (Erdil & Erkin, 2008). Instead, three levels of stability (Irving, 1994) have been defined in the SMT case, where ties are retained, that vary according to whether agents will agree to swap between choices they find equally acceptable. Under the weakest of these three definitions, which we assume in this paper, a stable matching can always be found by arbitrarily breaking the ties, resulting in an instance of SM.

If both ties and incomplete lists are introduced we obtain the Stable Marriage problem with Ties and Incomplete lists, or SMTI (Manlove, Irving, Iwama, Miyazaki, & Morita, 2002). In an instance of SMTI, stable matchings do not necessarily have the same size, and MAX-SMTI, the problem of finding a stable matching of maximum size, is \mathcal{NP} -hard (Manlove et al., 2002).

The Stable Roommates problem with Globally Ranked Pairs (SR-GRP) (Abraham, Levavi, Manlove, & O'Malley, 2008; Arkin et al., 2009) is a variant of the Stable Roommates problem involving ties and incomplete lists in which each pair of compatible agents $\{p, q\}$ has a weight $w(\{p, q\})$ assigned to their potential pairing, and the preference lists of each agent can be derived from these weights in the obvious manner: given two compatible pairs $\{p, q\}$ and $\{p, r\}$, p prefers q to r if and only if $w(\{p, q\}) > w(\{p, r\})$. This problem can be restricted to give the Stable Marriage problem with Ties, Incomplete lists, and Globally Ranked Pairs (SMTI-GRP) by splitting the agents into two sets as per the Stable Marriage problem.

In this work, we study one application of SMTI-GRP involving the pairing of children with adoptive families as coordinated by the British charity Coram.¹ Social workers determine a weight to be assigned to each child–family pair (c, f) , as a predicted measure of the suitability of placing c with f , giving an instance of SMTI-GRP. Currently Coram is using a clearing house system which pairs children and families at suitable specified intervals. Similar to the case for kidney exchange programmes (Roth, Sönmez, & Ünver, 2004), this allows for a more efficient pairing of children and families, at the cost of a slightly increased delay between entering the system and being paired. In such a system Coram has decided that the goal should be to find a stable matching that pairs as many children as possible and/or has maximum overall weight.² Moreover, Coram would like to ensure that the computed matching is viable in the long term. To this end, a lower bound, or threshold, on suitable weights is used to create refined instances of SMTI-GRP where child–family pairs with weights below the threshold are not allowed to be matched together. However, attempts to determine appropriate threshold values, as well as good weighting functions and suitable intervals between matching runs, have been hampered by the lack of tractable algorithms for finding maximum weight stable matchings for such instances. Indeed, in the SMTI-GRP setting, \mathcal{NP} -hardness holds for each of the problems of finding a maximum size stable matching (Abraham et al., 2008) and a maximum weight stable matching (Deligkas, Mertzios, & Spirakis, 2017).

Whilst SMTI is a one-to-one matching problem, in some applications one set of agents can be matched with more than one partner. The Hospitals/Residents (HR) problem (Gale & Shapley, 1962; Manlove, 2015) is a many-to-one extension of SMI that models the assignment of intending junior doctors (residents) to hospitals.

Table 1
Summary of matching problems.

Variant	Bipartite	Incompatible pairs	Ties	Weights	Capacity
SR	No	No	No	No	No
SR-GRP	No	Yes	Yes	Yes	No
SM	Yes	No	No	No	No
SMI	Yes	Yes	No	No	No
SMT	Yes	No	Yes	No	No
SMTI	Yes	Yes	Yes	No	No
SMTI-GRP	Yes	Yes	Yes	Yes	No
SMTI-SYM	Yes	Yes	Yes	Yes	No
HRT	Yes	Yes	Yes	No	Yes

Each doctor is to be assigned to at most one hospital, whilst each hospital may be assigned multiple doctors up to some given capacity. HR can be generalised to include ties in the preference lists, leading to the Hospitals/Residents with Ties (HRT), the many-to-one generalisation of SMTI. HRT has many applications: it models, for example, the assignment of medical graduates to Scottish hospitals as part of the Scottish Foundation Allocation Scheme (SFAS), which ran between 1999 and 2012. Since then, the UK has amalgamated all such schemes into the UK Foundation Programme, which handles the assignment of almost 8000 doctors to approximately 7000 positions across 20 Foundation Schools, each of which consists of multiple hospitals (UK Foundation Programme). In this setting a key aim is to find a stable matching of maximum size, which is an \mathcal{NP} -hard problem in view of the \mathcal{NP} -hardness of MAX-SMTI.

An overview of the differences between problems discussed in the paper is given in Table 1. The relationships between these problems are demonstrated in Fig. 1. In the diagram, an arrow from problem A to problem B indicates that problem B is a special case of problem A. For example, SMTI-SYM is the special case of SMTI-GRP in which preferences are symmetric.

1.2. Our contribution

In this paper, we have developed several new techniques that improve the performance of ILP models for instances of both SMTI and HRT. Our first contribution is to present two algorithms for preprocessing instances of SMTI with ties on both sides. Without such preprocessing, only 6 of 22 real-world instances from Coram could be solved within an hour per instance using state-of-the-art models from the literature. Our new preprocessing significantly improves this, finding solutions to 21 of the 22 instances in an average of 434 seconds. We also present new ILP models for SMTI and HRT. These use dummy variables to reduce the number of non-zero entries in their corresponding constraint matrices, which vastly increases the sparsity of the constraint matrix at the cost of additional variables. Further, we formulate different sets of constraints to model stability, including the use of redundant constraints to improve the continuous relaxations of our models. We test each of these individually, and these improvements together allow us to find solutions to all real-world instances in a mean runtime of less than 60 seconds. Turning to randomly-generated instances, the new models also solve all 30 random instances of SMTI that we generated with 50,000 agents on either side and preference lists of length 5 on one side, while existing state-of-the-art models could only solve 20. We extend our new ILP models to HRT, where we show a reduction in the mean runtime on existing real-world instances of HRT from SFAS, decreasing the average runtime from 144 seconds to only 3 seconds. We also generate 90 random instances that mimic the UK Foundation Programme (with about 7500 doctors and positions). Existing models solve 66 of these, while our new models solve 81.

¹ Coram—better chances for children since 1739, <https://www.coram.org.uk>.

² The child–family pairings in a computed stable matching are treated merely as suggestions that will be investigated further by social workers for suitability before any actual assignments are made.

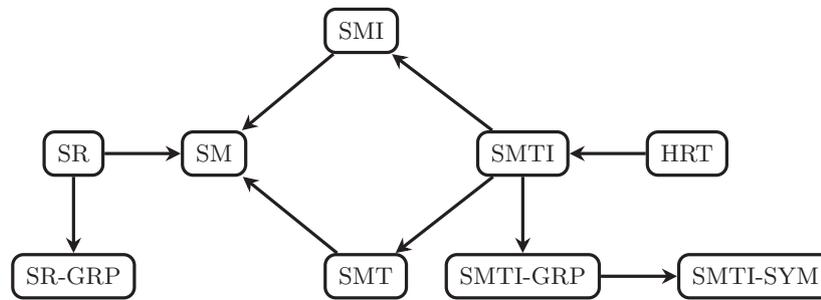


Fig. 1. Relationships between matching problems.

1.3. Related work

MAX-SMTI is known to be \mathcal{NP} -hard even if each tie occurs at the end of some agent's preference list, ties occur on one side only and each tie is of length two (Manlove et al., 2002). The special case of MAX-SMTI that asks if an instance of SMTI has a stable matching that matches every man and woman is also \mathcal{NP} -complete (Manlove et al., 2002), and this result holds even when preference lists have lengths of at most 3 and ties occur on one side only (Irving, Manlove, & O'Malley, 2009).

MAX-SMTI is also not approximable within a factor of $21/19$ (Halldórsson, Iwama, Miyazaki, & Yanagisawa, 2007) unless $\mathcal{P} = \mathcal{NP}$, even if preferences on one side are strictly ordered, and on the other side are either strictly ordered or a tie of length two. The best currently-known performance guarantee is $3/2$, achieved first in non-linear running time (McDermid, 2009) and later improved to linear time (Király, 2013; Paluch, 2014), although better guarantees can be achieved in certain restrictions (Iwama & Miyazaki, 2016). Király (2013) shows how to extend his $3/2$ -approximation algorithm for MAX-SMTI to HRT.

The Stable Marriage problem with Ties, Incomplete lists and Symmetric preferences (SMTI-SYM) is a restriction of SMTI-GRP such that (i) for each man–woman pair (u, v) , the rank of v in u 's list, i.e., the integer k such that v belongs to the k th tie in u 's list, is equal to the rank of u in v 's list, and (ii) the weight of (u, v) is precisely this integer k . Finding a maximum size stable matching in an instance of SMTI-SYM is \mathcal{NP} -hard, and therefore the same result holds for SMTI-GRP (Abraham et al., 2008). Given an instance of SMTI-GRP, if the goal is to find a matching that maximises the total weight rather than the total size, this problem is \mathcal{NP} -hard also (Deligkas et al., 2017).

Linear programming models for SM and SMI have been long studied, and stable matchings correspond exactly to extreme points of the solution polytopes of such models (Gusfield & Irving, 1989; Vande Vate, 1989). These formulations have been extended to give ILP models for finding maximum size stable matchings in instances of SMTI and HRT (Kwanashie, 2015; Kwanashie & Manlove, 2014). ILP models have also been given for a common extension of HR that allows doctors to apply as couples, typically so that both members can be matched to hospitals that are geographically close (Ágoston, Biró, & McBride, 2016; Drummond, Perault, & Bacchus, 2015; Hinder, 2015; Manlove, McBride, & Trimble, 2017; McBride, 2015). Other techniques in the field include constraint programming, which has been applied to SM and its variants (Gent, Irving, Manlove, Prosser, & Smith, 2001; Gent & Prosser, 2002; Manlove, O'Malley, Prosser, & Unsworth, 2005; O'Malley, 2007), and the use of SAT models and SAT solvers (Drummond et al., 2015; Gent & Prosser, 2002).

Diebold and Bichler (2017) performed a thorough experimental study of eight algorithms for HRT, giving a comparison of these algorithms when applied to real-world HRT instances derived from a course allocation system at the Technical University of Munich.

These datasets ranged in size from 18–733 students (the “doctors”) and 3–43 courses (the “hospitals”). The authors measured a number of attributes of the algorithms, including the sizes of the computed stable matchings. The methods that they considered included three exact algorithms for MAX-HRT based on the ILP model presented in Kwanashie and Manlove (2014).

Slaugh, Akan, Kesten, and Ünver (2016) described improvements they had made to the mechanism for matching children to adoptive families as utilised by the Pennsylvania Adoption Exchange. The process is semi-decentralised in that up to ten match attempts are made against families when each child arrives. By contrast, the more centralised process adopted by Coram involves a pool of children and families building up over time, leading to the use of a matching algorithm for the resulting two-sided matching problem.

For more details on the diverse variants of stable matching problems, we direct the reader to Manlove (2013) and for an economic overview of these problems we recommend (Roth & Sotomayor, 1990).

1.4. Layout of the paper

The rest of the paper is organised as follows. Section 2 defines the problems that are studied in this paper, and we introduce and discuss existing models for these in Section 3. This is followed by a theorem and two algorithms for preprocessing instances of SMTI in order to reduce instance sizes, in Section 4. Section 5 introduces our first new model, which reduces the number of non-zero elements in the constraint matrix through dummy variables. Further models are presented in Section 6 with new stability constraints. We demonstrate our new models and improvements experimentally in Section 7 and we provide some conclusions in Section 8.

2. Problem definitions

In this section we give formal definitions of the three key problems that we consider in this paper.

2.1. Stable Marriage with Ties and Incomplete Lists

An instance I of the *Stable Marriage problem with Ties and Incomplete lists (SMTI)* comprises a set C of n_1 children and a set F of n_2 families, where each child (respectively family) ranks a subset of the families (respectively children) in order of preference, possibly with ties. We say that a child $c \in C$ finds a family $f \in F$ acceptable if f belongs to c 's preference list, and we define acceptability for a family in a similar way. We assume that preference lists are *consistent*, that is, given a child–family pair $(c, f) \in C \times F$, c finds f acceptable if and only if f finds c acceptable. If c does find f acceptable then we call (c, f) an *acceptable pair*.

A *matching* M in I is a subset of acceptable pairs such that, for each agent $a \in C \cup F$, a appears in at most one pair in M . If a appears in a pair of M , we say that a is *matched*, otherwise a is *unmatched*. In the former case, $M(a)$ denotes a 's *partner* in M , that is,

if $(c, f) \in M$, then $M(c) = f$ and $M(f) = c$. We now define stability, which is the key condition that must be satisfied by a matching in I .

Definition 1. Let I be an instance of SMTI and let M be a matching in I . A child–family pair $(c, f) \in (C \times F) \setminus M$ is a *blocking pair* of M , or *blocks* M , if

1. (c, f) is an acceptable pair,
2. either c is unmatched in M or c prefers f to $M(c)$, and
3. either f is unmatched in M or f prefers c to $M(f)$.

M is said to be *stable* if it admits no blocking pair.

In SMTI, the goal is to find an arbitrary stable matching. We denote the problem of finding a maximum size stable matching, given an instance of SMTI, by MAX-SMTI.

2.2. Globally Ranked Pairs

An instance I of the *Stable Marriage problem with Ties, Incomplete lists and Globally-Ranked Pairs (SMTI-GRP)* comprises a set C of n_1 children, a set F of n_2 , a subset $X \subseteq C \times F$ of *acceptable* child–family pairs, and a weight function $w : X \rightarrow \mathbb{R}$.

The set of acceptable pairs and the weight function are used to define the SMTI instance I' corresponding to I as follows: for any two acceptable pairs $(c, f) \in X$ and $(c, f') \in X$, c prefers f to f' if $w(c, f) > w(c, f')$, and c is indifferent between f and f' if $w(c, f) = w(c, f')$. Preference lists of families are constructed in a similar fashion. A stable matching in I can then be defined by applying Definition 1 to I' .

Given a matching M in I , the *weight* of M , denoted by $w(M)$, is defined to be $\sum_{(c,f) \in M} w(c, f)$. The problem of finding a stable matching of maximum size is called MAX-SMTI-GRP, and the problem of finding a stable matching of maximum weight is called MAX-WT-SMTI-GRP.

Given an instance I of MAX-WT-SMTI-GRP, we can construct a refined instance I' of MAX-WT-SMTI-GRP from I by setting a threshold value t with the effect that the acceptable pairs in I' are precisely the acceptable pairs in I which have weight at least t . The effect of imposing different threshold values on I is of interest to Coram.

Example 1. Our first example demonstrates how different threshold values create instances of SMTI-GRP with differently sized maximum size stable matchings. Let $C = \{c_1, c_2, c_3\}$ be a set of children, $F = \{f_1, f_2, f_3\}$ be a set of families, and let the weight function w be defined by the following table:

	f_1	f_2	f_3
c_1	95	85	80
c_2	95	80	80
c_3	80	45	75

By taking $t = 0$ we obtain an instance of SMTI-GRP in which all pairs are acceptable. In this instance, $M_1 = \{(c_1, f_2), (c_2, f_1), (c_3, f_3)\}$ is the unique maximum weight stable matching and its weight is 255. However, if we take $t = 80$ and construct an instance of SMTI-GRP, then the only acceptable pair that involves c_3 is (c_3, f_1) and no stable matching can involve c_3 . The unique maximum weight stable matching is then $M_2 = \{(c_1, f_2), (c_2, f_1)\}$, which has a weight of 180.

Example 2. Our second instance of SMTI-GRP is intended to show that a maximum weight stable matching may be smaller in size than a maximum size stable matching. Let $C = \{c_1, c_2, c_3, c_4\}$, $F = \{f_1, f_2, f_3, f_4\}$,

$$X = \{(c_i, f_i) : 1 \leq i \leq 4\} \cup \{(c_{i+1}, f_i) : 1 \leq i \leq 3\},$$

and the weight function w be given by the following table:

	f_1	f_2	f_3	f_4
c_1	1	–	–	–
c_2	4	4	–	–
c_3	–	3	4	–
c_4	–	–	4	1

Let $M_1 = \{(c_i, f_i) : 1 \leq i \leq 4\}$ and $M_2 = \{(c_{i+1}, f_i) : 1 \leq i \leq 3\}$. It is easy to verify that M_1 and M_2 are both stable matchings. However $w(M_1) = 10$ and $w(M_2) = 11$, whereas $|M_1| = 4$ and $|M_2| = 3$.

2.3. Hospitals/Residents with Ties

An instance I of the *Hospitals/Residents problem with Ties (HRT)* comprises a set D of n_1 resident doctors and a set H of n_2 hospitals. Each doctor (respectively hospital) ranks a subset of the hospitals (respectively doctors) in order of preference, possibly with ties. Additionally, each hospital h has a *capacity* $c_h \in \mathbb{Z}^+$, meaning that h can be assigned at most c_h doctors, while each doctor is assigned to at most one hospital. The definitions of the terms *consistent* and *acceptable* are analogous to the SMTI case.

A *matching* M in I is a subset of acceptable pairs such that each doctor appears in at most one pair, and each hospital $h \in H$ appears in at most c_h pairs. Given a doctor $d \in D$, the terms *matched* and *unmatched*, and the notation $M(d)$, are defined as in the SMTI case. Given a hospital $h \in H$, we let $M(h) = \{d \in D : (d, h) \in M\}$. We say that h is *full* or *undersubscribed* in M if $|M(h)| = c_h$ or $|M(h)| < c_h$, respectively. We next define stability by extending Definition 1 to the HRT case.

Definition 2. Let I be an instance of HRT and let M be a matching in I . A doctor–hospital pair $(d, h) \in (D \times H) \setminus M$ is a *blocking pair* of M , or *blocks* M , if

1. (d, h) is an acceptable pair,
2. either d is unmatched in M or d prefers h to $M(d)$, and
3. either h is undersubscribed in M or h prefers d to some member of $M(h)$.

M is said to be *stable* if it admits no blocking pair.

As in the SMTI case, the problem of finding a maximum size stable matching, given an instance of HRT, is denoted MAX-HRT.

While the definition for HRT does allow an arbitrary number of preferences to be expressed by any doctor, in reality doctors' preference lists are often short: for example in the SFAS application until 2009, every doctor's list was of length 6.

3. Existing formulations

The first mathematical models for SM were proposed in the late 1980s independently by [Gusfield and Irving \(1989\)](#) and by [Vande Vate \(1989\)](#). [Rothblum \(1992\)](#) extended Vande Vate's model to SMI. In the following, we show how to extend Rothblum's model to formulate both MAX-SMTI and MAX-HRT, as was done previously by [Kwanashie and Manlove \(2014\)](#). These existing models for MAX-SMTI and MAX-HRT are described here as they will be extended in later sections.

3.1. Mathematical model for MAX-SMTI

Based on our Coram application, we will adopt the terminology from that context when presenting models for MAX-SMTI. When reasoning about models, we will use i and j to represent a child and family, rather than c and f , respectively, as i and j are by convention more typically used as subscript variables. Let us consider the following notation:

- $F(i)$ is the set of families acceptable for child i ($i = 1, \dots, n_1$).
- $C(j)$ is the set of children acceptable for family j ($j = 1, \dots, n_2$).
- $r_j^c(i)$ is the rank of family j for child i , defined as the integer k such that j belongs to the k th most-preferred tie in i 's list ($i = 1, \dots, n_1, j \in F(i)$). The smaller the value of $r_j^c(i)$, the better family j is ranked for child i .
- $r_i^f(j)$ is the rank of child i for family j , defined as the integer k such that i belongs to the k th most-preferred tie in j 's list ($j = 1, \dots, n_2, i \in C(j)$). The smaller the value of $r_i^f(j)$, the better child i is ranked for family j .
- $F_j^{\leq}(i)$ is the set of families that child i ranks at the same level or better than family j , that is, $F_j^{\leq}(i) = \{j' \in F : r_{j'}^c(i) \leq r_j^c(i)\}$ ($i = 1, \dots, n_1, j \in F(i)$).
- $C_i^{\leq}(j)$ is the set of children that family j ranks at the same level or better than child i , that is, $C_i^{\leq}(j) = \{i' \in C : r_{i'}^f(j) \leq r_i^f(j)\}$ ($j = 1, \dots, n_2, i \in C(j)$).

By introducing binary decision variables x_{ij} that take value 1 if child i is matched with family j , and 0 otherwise ($i = 1, \dots, n_1, j \in F(i)$), MAX-SMTI can be modelled as follows:

$$\max \sum_{i=1}^{n_1} \sum_{j \in F(i)} x_{ij} \tag{1}$$

$$\text{s.t.} \sum_{j \in F(i)} x_{ij} \leq 1, \quad i = 1, \dots, n_1, \tag{2}$$

$$\sum_{i \in C(j)} x_{ij} \leq 1, \quad j = 1, \dots, n_2, \tag{3}$$

$$1 - \sum_{q \in F_j^{\leq}(i)} x_{iq} \leq \sum_{p \in C_i^{\leq}(j)} x_{pj}, \quad i = 1, \dots, n_1, j \in F(i), \tag{4}$$

$$x_{ij} \in \{0, 1\}, \quad i = 1, \dots, n_1, j \in F(i). \tag{5}$$

The objective function (1) maximises the number of children assigned. If instead, one wants to maximise the score of the children assigned (as in MAX-WT-SMTI-GRP), it is enough to use $\sum_{i=1}^{n_1} \sum_{j \in F(i)} w_{ij} x_{ij}$ in the objective function. Constraints (2) ensure that each child is matched with at most one family and constraints (3) impose that each family is matched with at most one child. Finally, constraints (4) enforce stability by ruling out the existence of any blocking pair. More specifically, they ensure that if child i is not matched with family j or any other family they rank at the same level or better than j (i.e., $\sum_{q \in F_j^{\leq}(i)} x_{iq} = 0$), then family j is matched with a child it ranks at the same level or better than i (i.e., $\sum_{p \in C_i^{\leq}(j)} x_{pj} \geq 1$).

3.2. Mathematical model for MAX-HRT

An adaptation of model (1)–(5) for MAX-HRT was proposed in Kwanashie and Manlove (2014). It uses the same notation that was used for MAX-SMTI except that:

- The term “family” is replaced by “hospital” and $F(i)$, $r_j^c(j)$, and $F_j^{\leq}(i)$ are changed into $H(i)$, $r_i^h(j)$, and $H_j^{\leq}(i)$, respectively.
- The term “child” is replaced by “doctor” and $C(j)$, $r_j^c(i)$, and $C_i^{\leq}(j)$ are changed to $D(j)$, $r_j^d(i)$, and $D_i^{\leq}(j)$, respectively.
- The capacity of hospital j ($j = 1, \dots, n_2$) is denoted by c_j .

By introducing binary decision variables x_{ij} that take value 1 if doctor i is assigned to hospital j , and 0 otherwise ($i =$

$1, \dots, n_1, j \in H(i)$), MAX-HRT can be modelled as follows:

$$\max \sum_{i=1}^{n_1} \sum_{j \in H(i)} x_{ij} \tag{6}$$

$$\text{s.t.} \sum_{j \in H(i)} x_{ij} \leq 1, \quad i = 1, \dots, n_1, \tag{7}$$

$$\sum_{i \in D(j)} x_{ij} \leq c_j, \quad j = 1, \dots, n_2, \tag{8}$$

$$c_j \left(1 - \sum_{q \in H_j^{\leq}(i)} x_{iq} \right) \leq \sum_{p \in D_i^{\leq}(j)} x_{pj}, \quad i = 1, \dots, n_1, j \in H(i), \tag{9}$$

$$x_{ij} \in \{0, 1\}, \quad i = 1, \dots, n_1, j \in H(i). \tag{10}$$

While the meaning of the objective function and constraints (7) remains the same, constraints (8) ensure now that each hospital does not exceed its capacity. Constraints (9) are the adaptation of the stability constraints (4) when capacity is considered. More specifically, they ensure that if doctor i was not assigned to hospital j or any other hospital they rank at the same level or higher than j (i.e., $\sum_{q \in H_j^{\leq}(i)} x_{iq} = 0$), then hospital j has filled its capacity with doctors it ranks at the same level or higher than i (i.e., $\sum_{p \in D_i^{\leq}(j)} x_{pj} \geq c_j$).

3.3. Discussion on the models

Although the model for SM was proposed almost thirty years ago, the computational behaviour of its extension to MAX-SMTI and MAX-WT-SMTI-GRP (i.e., in one-to-one instances specifically) has never been studied, to the best of our knowledge. However, we mention that our direct implementation of (1)–(5) on real-world MAX-WT-SMTI-GRP instances involving 500 children, 1000 families, and a large list of preferences cannot be solved by state-of-the-art solvers within hours. Indeed, the model becomes too difficult as it requires up to 500,000 stability constraints, each of them including $|F_j^{\leq}(i)| + |C_i^{\leq}(j)|$ nonzero elements (i.e., up to 1500).

Regarding MAX-HRT, computational experiments with (6)–(10) applied to real-world and randomly generated instances have been carried out previously (Diebold & Bichler, 2017; Kwanashie, 2015; Kwanashie & Manlove, 2014). Kwanashie (2015) observed a significant increase in terms of average running time when the number of doctors goes above 400. As our objective is to solve instances of the magnitude of the UK Foundation Programme application (involving almost 8000 doctors and 500 hospitals), the model in its current form is not suitable.

In the next sections, we introduce various techniques aimed at reducing the size of the two models and strengthening their continuous relaxation.

4. Preprocessing SMTI with ties on both sides

It is quite common in combinatorial optimisation to use some simple analysis to fix the optimal value of a subset of variables and, thus, reduce the problem size. This is particularly useful for stable matching problems as one variable, one stability constraint, and up to $n_1 + n_2$ non-zero elements are associated with each acceptable pair. Two procedures, “Hospitals-offer” and “Residents-apply”, have been proposed for removing acceptable pairs that cannot be part of any stable matching for HRT when ties only occur in hospitals’ preference lists (Irving & Manlove, 2009).

When ties can belong to the preference lists of both sets of agents, a reduction technique is known for the special case of SMTI

in which preference lists on one side are of length at most two (Irving et al., 2009). However the aforementioned preprocessing algorithms are not applicable to arbitrary instances of SMTI. In this section we introduce a new sufficient condition to find a set of acceptable pairs that cannot be part of any stable matching for SMTI. We then propose two greedy algorithms to detect such pairs which can then be removed from the instance without affecting any stable matching. Our technique is based on the following result.

Theorem 1. *Let I be an instance of SMTI. Given a child i and a set of families \mathcal{F} such that for every family $j \in \mathcal{F}$, (i, j) is an acceptable pair, let \mathcal{C} be the set of children that at least one family in \mathcal{F} ranks at the same level or better than i , i.e., $\mathcal{C} = \bigcup_{j \in \mathcal{F}} \{C_j^{\leq}(i)\}$. If $|\mathcal{F}| \geq |\mathcal{C}|$, then in any stable matching M , child i will be matched with a family j' such that $r_{j'}^c(i) \leq \max_{j \in \mathcal{F}} \{r_j^c(i)\}$.*

Proof. Assume for a contradiction that M is a stable matching in I in which child i is matched with a family j' with $r_{j'}^c(i) > \max_{j \in \mathcal{F}} r_j^c(i)$ or is unmatched. Since $|\mathcal{F}| \geq |\mathcal{C}| > |\mathcal{C} \setminus \{i\}|$, at least one family $j'' \in \mathcal{F}$ must be matched with some child $i' \notin \mathcal{C}$ or be unmatched. Then either i is unmatched or prefers j'' to j' , and either j'' is unmatched or prefers i to i' . In all cases (i, j'') blocks M , which is a contradiction. \square

There is no obvious efficient way to find, for each child, the set \mathcal{F} that removes the largest number of acceptable pairs from an instance of SMTI. Instead we present two polynomial-time algorithms to find sets that allow a significant number of acceptable pairs to be removed. Algorithm 1, “first-rank-family”, considers the

Algorithm 1 First-rank-family.

```

1: Input: An instance of SMTI with children  $C$  and families  $F$ 
2: Output: A set  $\mathcal{R}$  containing pairs  $(i, j')$  that cannot be part of any stable matching
3: for each  $\mathcal{C} \in \mathcal{P}(C)$  do  $\triangleright$  for each subset of children in the powerset  $\mathcal{P}(C)$ 
4:    $\mathcal{M}_{\mathcal{C}} \leftarrow \emptyset$ 
5: end for
6:  $\mathcal{R} \leftarrow \emptyset$ 
7: for each family  $j \in F$  do
8:    $\mathcal{C} \leftarrow \{i \in C(j) : r_i^f(j) = 1\}$   $\triangleright$  the set of children family  $j$  considers equally best
9:    $\mathcal{M}_{\mathcal{C}} \leftarrow \mathcal{M}_{\mathcal{C}} \cup \{j\}$ 
10: end for
11: for each  $\mathcal{C} \in \mathcal{P}(C)$  do
12:    $\mathcal{F} \leftarrow \mathcal{M}_{\mathcal{C}}$ 
13:   if  $|\mathcal{F}| \geq |\mathcal{C}|$  then
14:     for each  $i \in \mathcal{C}$  do
15:       for each  $j' \in F(i)$  with  $r_{j'}^c(i) > \max_{j \in \mathcal{F}} \{r_j^c(i)\}$  do
16:          $\mathcal{R} \leftarrow \mathcal{R} \cup \{(i, j')\}$ 
17:       end for
18:     end for
19:   end if
20: end for
21: return  $\mathcal{R}$ 

```

first rank of children for each family j , i.e., the children that j thinks are the most desirable. Algorithm 2, “full-child-preferences”, completely analyses the preference lists of the children to find reductions. Note that each of these algorithms can also be applied to the preferences of the other set of agents by symmetry to obtain “first-rank-child” and “full-family-preferences”, and that they may each be applied iteratively until no more reductions are possible.

After initialisation (lines 3–6), Algorithm 1 considers each family j in turn, determining the set of children \mathcal{C} that family j ranks as (equally) most desirable (line 8) and storing this fact (line 9). Once

Algorithm 2 Full-child-preferences.

```

1: Input: An instance of SMTI with children  $C$  and families  $F$ 
2: Output: A set  $\mathcal{R}$  containing pairs  $(i, j')$  that cannot be part of any stable matching
3: for each child  $i \in C$  do
4:    $\mathcal{F} \leftarrow \emptyset$ 
5:    $\mathcal{C} \leftarrow \emptyset$ 
6:   for each  $j \in F(i)$  do  $\triangleright$  for each family in descending order of preference
7:      $\mathcal{F} \leftarrow \mathcal{F} \cup \{j\}$ 
8:      $\mathcal{C} \leftarrow \mathcal{C} \cup \{i' \in C(j) : r_{i'}^f(j) \leq r_i^f(j)\}$ 
9:     if  $|\mathcal{F}| \geq |\mathcal{C}|$  then
10:       for each  $j' \in F(i)$  with  $r_{j'}^c(i) > \max_{j \in \mathcal{F}} \{r_j^c(i)\}$  do
11:          $\mathcal{R} \leftarrow \mathcal{R} \cup \{(i, j')\}$ 
12:       end for
13:     break
14:   end if
15: end for
16: end for
17: return  $\mathcal{R}$ 

```

this has been recorded, the algorithm searches through all these stored sets (line 11) to find sets of children \mathcal{C} and the set of families \mathcal{F} which all consider the set \mathcal{C} as their (equally) best choice. If the set of families \mathcal{F} is at least as big as the set of children \mathcal{C} (line 13) then for each child $i \in \mathcal{C}$ and each family $j' \in F(i)$ ranked worse than the worst family in \mathcal{F} , we add the pair (i, j') to our reduction set \mathcal{R} (lines 14–16).

As written, Algorithm 1 requires $\mathcal{O}(2^{n_1} n_1 n_2)$ operations, as we must iterate over each possible subset of children (in both lines 3 and 11). However, if we only explicitly store the subsets \mathcal{C} and $\mathcal{M}_{\mathcal{C}}$ generated by lines 7–10, we will obtain at most n_2 subsets \mathcal{C} and at most n_2 subsets $\mathcal{M}_{\mathcal{C}}$. To only store these specific subsets, we need to quickly look up whether such a set \mathcal{C} exists, and create it if it does not, before adding a family j to $\mathcal{M}_{\mathcal{C}}$. A hashmap is a suitable data structure for carrying out these operations, and will reduce the overall complexity to $\mathcal{O}(n_1 n_2^2)$.

Algorithm 2 incrementally builds up the sets \mathcal{F} and \mathcal{C} for each child i . To build \mathcal{F} , we simply add each family j from the preference list of i in order from most preferable to least (lines 6 and 7), considering agents within ties in increasing indicial order. At each step, when we have added j , we then add to \mathcal{C} all children that j finds at least as preferable as i (line 8). By construction these satisfy Theorem 1. Thus, if \mathcal{F} is large enough compared to \mathcal{C} , we add to our reduction all the pairs (i, j') where $j' \in F(i)$ are the families ranked worse than the worst family in \mathcal{F} (lines 9–11). Algorithm 2 requires $\mathcal{O}(n_1 n_2 (n_1 + n_2))$ steps as the outer (respectively middle and inner) **for each** loop is executed $\mathcal{O}(n_1)$ (respectively $\mathcal{O}(n_2)$ and $\mathcal{O}(n_2)$) times, and line 8 requires $\mathcal{O}(n_1)$ time.

We note that: (i) this preprocessing is more powerful when the number of ranks (i.e., groups of tied elements) is high and when there are only a few agents in each rank, and (ii) rather than adding families in descending order of preference, more sophisticated heuristics could find a larger number of reductions at the cost of a higher time complexity. However, it is worth mentioning that our greedy approach works particularly well when there is a strong correlation between the scores obtained by a given agent among the other agents, e.g., if a family is ranked first for a given child, it also tends to be ranked highly by other children, which is the case in our application. We show in Section 7 that the greedy approaches given by Algorithms 1 and 2 can significantly reduce running times for our SMTI-GRP instances. We also remark that we did not try to extend Algorithms 1 and 2 to HRT instances with ties on both sides, as our practical application involving SFAS instances

allows ties on one side only, and in such a setting we may apply Algorithms “Hospitals-offer” and “Residents-apply” from Irving and Manlove (2009).

We conclude this section with an example of the application of Algorithms 1 and 2.

Example 3. Let us consider an SMTI instance with 5 families and 4 children with the following preference lists:

$c_1 : (f_1 f_2 f_3) f_4$	$f_1 : (c_1 c_3) c_4$
$c_2 : (f_2 f_3 f_4) f_5$	$f_2 : (c_1 c_2) c_4$
$c_3 : (f_1 f_3 f_4)$	$f_3 : (c_2 c_3) c_1$
$c_4 : (f_1 f_2 f_4)$	$f_4 : (c_1 c_2) (c_3 c_4)$
	$f_5 : c_2.$

In this example, child 1 prefers to be matched equally with families 1, 2, and 3. If his first choice is not granted, then child 1 prefers to be matched with family 4.

We start by running “first-rank-child”, but we see that no two children share the same common set of families as their first preference, so no acceptable pair is removed. We then run “first-rank-family”, which highlights that both f_2 and f_4 have the same pair of children as their equally-first choice (c_1 and c_2). This tells us that children c_1 and c_2 will never be matched with a family that they prefer less than both f_2 and f_4 . Therefore, there is no need for c_2 to ever consider f_5 . This leaves the following preferences.

$c_1 : (f_1 f_2 f_3) f_4$	$f_1 : (c_1 c_3) c_4$
$c_2 : (f_2 f_3 f_4)$	$f_2 : (c_1 c_2) c_4$
$c_3 : (f_1 f_3 f_4)$	$f_3 : (c_2 c_3) c_1$
$c_4 : (f_1 f_2 f_4)$	$f_4 : (c_1 c_2) (c_3 c_4).$

As the instance was reduced, we could now re-run “first-rank-child” to see if any further reductions are to be found. However, no more reductions will be found, and so we move on to “full-child-preferences” and “full-family-preferences”. We demonstrate the former on child c_1 to obtain the following sequence of sets \mathcal{F} and \mathcal{C} :

$\mathcal{F} = \{f_1\}$	$\mathcal{C} = \{c_1, c_3\}$
$\mathcal{F} = \{f_1, f_2\}$	$\mathcal{C} = \{c_1, c_2, c_3\}$
$\mathcal{F} = \{f_1, f_2, f_3\}$	$\mathcal{C} = \{c_1, c_2, c_3\}.$

As $|\mathcal{F}| \geq |\mathcal{C}|$, we know that c_1 cannot be matched with a family that c_1 would rank as worse than the worst family in \mathcal{F} . This means that c_1 will never consider f_4 , so the acceptable pair (c_1, f_4) can be removed, leaving the following reduced instance.

$c_1 : (f_1 f_2 f_3)$	$f_1 : (c_1 c_3) c_4$
$c_2 : (f_2 f_3 f_4)$	$f_2 : (c_1 c_2) c_4$
$c_3 : (f_1 f_3 f_4)$	$f_3 : (c_2 c_3) c_1$
$c_4 : (f_1 f_2 f_4)$	$f_4 : c_2 (c_3 c_4).$

Since we did reduce the instance, it is possible that re-running one of the other algorithms might reduce the instance even further, but in this particular instance no more reductions can be found.

5. Reducing the number of non-zero elements

Even if the reduction procedures previously described remove a significant number of acceptable pairs, the models involved in real-world instances remain too large to be solved by state-of-the-art ILP solvers. There are $O(n_1 n_2)$ constraints and variables and up to $O(n_1 n_2 (n_1 + n_2))$ non-zero elements, depending on the length of the agents’ preference lists. In this section, we propose an alternative formulation for MAX-SMTI that uses dummy variables to

keep track of the children’s and families’ assignments at each rank so that the overall number of non-zero elements is reduced. Let us consider the following additional notation:

- $g^c(i)$ is the number of distinct ranks (or ties) for child i ($i = 1, \dots, n_1$).
- $g^f(j)$ is the number of distinct ranks for family j ($j = 1, \dots, n_2$).
- $F_k^c(i)$ is the set of families acceptable for child i ($i = 1, \dots, n_1$) with rank k ($k = 1, \dots, g^c(i)$).
- $C_k^c(j)$ is the set of children acceptable for family j ($j = 1, \dots, n_2$) with rank k ($k = 1, \dots, g^f(j)$).

In addition, we introduce the dummy binary decision variables y_{ik}^c (respectively, y_{jk}^f) that take value 1 if child i (respectively, family j) is matched with a family (respectively, a child) of rank at most k , and 0 otherwise ($i = 1, \dots, n_1, k = 1, \dots, g^c(i)$) (respectively, $j = 1, \dots, n_2, k = 1, \dots, g^f(j)$). Variables y_{ik}^c and y_{jk}^f can be seen as a replacement of the summations of x_{iq} and x_{pj} over the sets $F_j^c(i)$ and $C_i^c(j)$. These variables have certain similarities with the cut-off scores for the college admission problem (Ágoston et al., 2016) and the radius formulation for the p -median problem (García, Labbé, & Marín, 2011).

The new formulation for MAX-SMTI is:

$$\max \sum_{i=1}^{n_1} y_{i, g^c(i)}^c \tag{11}$$

$$\text{s.t.} \sum_{j \in F_k^c(i)} x_{ij} = y_{i1}^c, \quad i = 1, \dots, n_1, \tag{12}$$

$$\sum_{j \in F_k^c(i)} x_{ij} + y_{i, k-1}^c = y_{ik}^c, \quad i = 1, \dots, n_1, k = 2, \dots, g^c(i), \tag{13}$$

$$\sum_{i \in C_1^c(j)} x_{ij} = y_{j1}^f, \quad j = 1, \dots, n_2, \tag{14}$$

$$\sum_{i \in C_k^c(j)} x_{ij} + y_{j, k-1}^f = y_{jk}^f, \quad j = 1, \dots, n_2, k = 2, \dots, g^f(j), \tag{15}$$

$$1 - y_{ik}^c \leq y_{j, r_i^f(j)}^f, \quad i = 1, \dots, n_1, k = 1, \dots, g^c(i), j \in F_k^c(i), \tag{16}$$

$$x_{i,j} \in \{0, 1\}, \quad i = 1, \dots, n_1, j \in F(i), \tag{17}$$

$$y_{ik}^c \in \{0, 1\}, \quad i = 1, \dots, n_1, k = 1, \dots, g^c(i), \tag{18}$$

$$y_{jk}^f \in \{0, 1\}, \quad j = 1, \dots, n_2, k = 1, \dots, g^f(j). \tag{19}$$

The objective function (11) now uses the last y_{ik}^c variable for each child (i.e., the one associated with its last rank) as an indicator of whether the child is assigned to a family. First, we note that even if (11) uses fewer non-zero elements than (1), both objective functions are equivalent. Second, the version of (1) that considers the weight of each pair should be used to solve MAX-WT-SMTI-GRP as (11) cannot be adapted for the problem. Constraints (12)–(15) maintain the coherence of variables y_{ik}^c and y_{jk}^f . Constraints (16) ensure the stability of the matching by using the new variables: if child i was not matched with a family of rank k or better (i.e., $1 - y_{ik}^c = 1$), that means that all families that child i ranks at level k were already matched with a child of better or equal rank (i.e., $y_{j, r_i^f(j)}^f \geq 1 \forall j \in F_k^c(i)$). Finally, by imposing binary values, constraints (18) and (19) prevent any child or family from being matched more than once. Note that the model would also be

valid if variables y_{ik}^c and y_{jk}^f were defined as continuous. However, preliminary experiments showed that it was not beneficial to do so.

Model (11)–(19) requires $O(\sum_{i=1}^{n_1} g^c(i) + \sum_{j=1}^{n_2} g^f(j))$ additional variables. It still uses $O(n_1 n_2)$ stability constraints, but they now involve only two variables, which reduces the overall size of the model.

By adopting similar notation for MAX-HRT, where $g^d(i)$ is the number of ranks (or ties) for doctor i ($i = 1, \dots, n_1$), $g^h(j)$ is the number of ranks (ties) for hospital j ($j = 1, \dots, n_2$), y_{ik}^d is a binary decision variable that takes the value 1 if and only if doctor i is assigned to a hospital of rank at most k , and y_{jk}^h is an integer decision variables indicating how many doctors of rank at most k are assigned to hospital j , MAX-HRT becomes:

$$\max \sum_{i=1}^{n_1} y_{i, g^d(i)}^d \tag{20}$$

$$\text{s.t. } y_{j, g^h(j)}^h \leq c_j, \quad j = 1, \dots, n_2, \tag{21}$$

$$c_j(1 - y_{ik}^d) \leq y_{j, r_i^h(j)}^h, \quad i=1, \dots, n_1, \quad k = 1, \dots, g^d(i), \quad j \in H_k^-(i), \tag{22}$$

$$y_{jk}^h \in \mathbb{Z}^+, \quad j = 1, \dots, n_2, \quad k = 1, \dots, g^h(j), \tag{23}$$

(12)–(15), (17), (18),

where (12)–(15) and (17), (18) are appropriately modified to follow HRT notation.

6. Alternative stability constraints

While dummy variables reduce the number of non-zero elements involved in the stability constraints, we introduce in this section some additional techniques that influence the number of stability constraints and the quality of the continuous relaxations of the models. It is well-known that the performance of an integer model depends not only on its size, but also on its linear relaxation. It was shown in the literature that for several problems (see, e.g., the Bin Packing Problem, Delorme, Iori, & Martello, 2016; or the Resource-Constrained Project Scheduling Problem, Koné, Artigues, Lopez, & Mongeau, 2011), it may be beneficial to use larger models if they have a better continuous relaxation (i.e., closer to the optimal solution).

6.1. Reduced stability constraints for MAX-SMTI

6.1.1. Constraint merging

Model (11)–(19) can be further reduced by merging, for a given child, all stability constraints with the same rank. Constraints (16) now become

$$|F_k^-(i)|(1 - y_{ik}^c) \leq \sum_{j \in F_k^-(i)} y_{j, r_i^f(j)}^f, \quad i = 1, \dots, n_1, \quad k = 1, \dots, g^c(i). \tag{24}$$

This transformation reduces the size of the model, as it uses only $O(\sum_{i=1}^{n_1} g^c(i))$ stability constraints. However, as will be shown in the computational experiments section, it also leads to a deterioration of the continuous relaxation bound. We note that the reduction in terms of size with respect to model (11)–(19) is more significant when the number of ranks (i.e., tie groups) is low.

6.1.2. Double stability constraints

To compensate for the degradation of the continuous relaxation caused by the constraint merging, it is possible to use the additional stability constraints

$$|C_k^-(j)|(1 - y_{jk}^f) \leq \sum_{i \in C_k^-(j)} y_{i, r_j^c(i)}^c, \quad j = 1, \dots, n_2, \quad k = 1, \dots, g^f(j). \tag{25}$$

These constraints can be seen as the counterparts of (24) when the merging is performed on the families instead of the children. These additional constraints improve the quality of the continuous relaxation with respect to the model that uses only (24). Overall, we observe a tradeoff between the number of stability constraints used in the model and the quality of the bound obtained by the continuous relaxation.

6.2. New stability constraints for MAX-HRT

For MAX-HRT, merging constraint (22) is not useful if there are no ties on the doctors' side (i.e., if $|H_k^-(i)| = 1$, $i = 1, \dots, n_1$, $k = 1, \dots, g^d(i)$). As our practical case allows ties on the hospitals' side only, it is not an improvement we explored. In this section, we propose instead an enriched formulation for MAX-HRT that allows us to define a second set of stability constraints. We introduce new binary decision variables z_{jk} that take value 1 if hospital j has filled entirely its capacity with doctors of rank at most $k - 1$, and 0 otherwise ($j = 1, \dots, n_2$, $k = 1, \dots, g^h(j) + 1$). An additional set of stability constraints for MAX-HRT is:

$$x_{ij} \leq 1 - z_{j, r_i^h(j)}, \quad i = 1, \dots, n_1, \quad j \in H(i), \tag{26}$$

$$z_{jk} \geq z_{j, k-1}, \quad j = 1, \dots, n_2, \quad k = 2, \dots, g^h(j) + 1, \tag{27}$$

$$1 - z_{jk} \leq y_{i, r_j^d(i)}^d, \quad j = 1, \dots, n_2, \quad k = 2, \dots, g^h(j) + 1, \quad i \in D_{k-1}^-(j), \tag{28}$$

$$c_j z_{j, g^h(j)+1} \leq y_{j, g^h(j)}^h, \quad j = 1, \dots, n_2, \tag{29}$$

$$z_{jk} \in \{0, 1\}, \quad j = 1, \dots, n_2, \quad k = 1, \dots, g^h(j) + 1. \tag{30}$$

Constraints (26) ensure that a doctor can only be assigned to a hospital that is not already filled by doctors that the hospital strictly prefers. Constraints (27) ensure that, if a hospital is full for doctors of rank at most $k - 1$, then it is also full for doctors of rank at most k . Constraints (28) ensure that the matching is stable by ruling out the existence of any blocking pair. More specifically, (28) ensure that, if a hospital j has space for doctors of rank k (i.e., $z_{jk} = 0$), then all doctors i of the hospital with rank $k - 1$ were already accepted in j or in a hospital they consider equal or better than j (i.e., $y_{i, r_j^d(i)}^d = 1$). Finally, constraints (29) ensure that if the hospital is full ($z_{j, g^h(j)+1} = 1$), then it has c_j doctors assigned to it ($y_{j, g^h(j)}^h \geq c_j$).

As ties occur on the hospital side, constraint merging can be applied to (28) to obtain:

$$|D_{k-1}^-(j)|(1 - z_{jk}) \leq \sum_{i \in D_{k-1}^-(j)} y_{i, r_j^d(i)}^d, \quad j = 1, \dots, n_2, \quad k = 2, \dots, g^h(j) + 1. \tag{31}$$

Note that both sets of stability constraints (22) and (26)–(30) can be used at the same time. Moreover, stability constraints

$$c_j z_{jk} \leq y_{j, k-1}^h, \quad j = 1, \dots, n_2, \quad k = 2, \dots, g^h(j) + 1, \tag{32}$$

stating that if a given hospital j has no room for doctors at rank k (i.e., $z_{jk} = 1$), then it has already selected c_j doctors of rank at most $k - 1$ (i.e., $y_{j, k-1}^h \geq c_j$), could be used to replace (22). Indeed, let us consider a doctor i , a hospital j , and their respective ranks $r_j^d(i) = m$ and $r_i^h(j) = p$. By (28), we know that

$$1 - z_{j, p+1} \leq y_{im}^d, \tag{33}$$

Table 2
Summary of the experiments.

Problem	Table	Dataset		Purpose
		#Inst.	Source	
MAX-WT-SMTI-GRP	3	22	Coram	Impact of dummy var. and alternative stability cons. Impact of preprocessing, warm start, and priorities Models' limits
	4	22	Coram	
	5	220	Randomly generated	
MAX-SMTI	6–8	270	Randomly generated	Impact of dummy var. and alternative stability cons.
MAX-HRT non master	9	3	SFAS	Impact of dummy var. and alternative stability cons.
	10	700	Randomly generated	Impact of dummy var. and alternative stability cons.
	11	150	Randomly generated	Models' limits
	12	60	Randomly generated	Impact of preprocessing, warm start, and priorities
MAX-HRT master	13	450	Randomly generated	Difference between the initial and the best model

which can be rewritten as

$$c_j(1 - y_{im}^d) \leq c_j z_{j,p+1}. \quad (34)$$

This can be completed by (32) to obtain

$$c_j(1 - y_{im}^d) \leq c_j z_{j,p+1} \leq y_{jp}^h, \quad (35)$$

which leads to constraints (22) being redundant. Notice that while we count $O(n_1 n_2)$ stability constraints (22), only $O(\sum_{j=1}^{n_2} g^h(j))$ are required for (32).

7. Computational experiments

We report in this section the outcome of extensive computational experiments aimed at testing the effectiveness of the proposed improvements for MAX-WT-SMTI-GRP, MAX-SMTI, and MAX-HRT. All algorithms were coded in C++, and Gurobi 7.5.2 was used to solve the ILP models. The implemented software is downloadable from the online repository <https://dx.doi.org/10.5281/zenodo.2538150>. The experiments were run on an Intel Xeon E5-2687W v3, 3.10 gigahertz with 512 gigabytes of memory, running under Linux 4.13.0. Each instance was run using a single core and had a total time limit (comprising model creation time and solution time) of 3600 seconds per problem instance. The instances that were randomly generated are downloadable from the online repository <https://dx.doi.org/10.5525/gla.researchdata.664>.

For each problem, a first set of experiments determined what combination of the improvements proposed in Sections 5 and 6 is the most effective. At a second stage, we reran a subset of these combinations to evaluate the impact of other features (such as preprocessing, branching priorities, and warm start). Experimental evaluations of the algorithms for MAX-WT-SMTI-GRP, MAX-SMTI, and MAX-HRT are presented in Sections 7.1, 7.2, and 7.3 respectively. A summary of the experiments reported in this section is presented in Table 2. For each experiment, the table identifies the problem solved, the subsequent tables containing the results, the dataset used (number of instances and source), and the experimental objectives. We show that dummy variables are particularly useful for MAX-WT-SMTI-GRP, constraint merging is beneficial for MAX-SMTI, and the new set of stability constraints substantially improves the performance of MAX-HRT. A more detailed discussion of these results can be found in Section 7.4

7.1. MAX-WT-SMTI-GRP

7.1.1. Real-world instances

We were provided with a sample set of data representing 550 children and 894 families, which included a weight (determined by Coram) for every child–family pair. Since most of the weights vary between 80 and 100, for each integer threshold in

{0, 80, 81, ..., 100}, we created an instance of SMTI-GRP as described in Section 2, resulting in 22 instances. When the threshold is set to 0, all child–family pairs are acceptable, and so technically we have complete rather than incomplete preference lists. Our models are still applicable to such instances.

The sample dataset contained a significant number of ties. One way to measure the density of ties in an SMTI or HRT instance is now described. Given an instance of SMI (i.e., with no ties), an instance of SMTI can be created as follows. For each set of agents pick a *tie density* t_d with $0 \leq t_d \leq 1$ (i.e., so the children and families may have distinct tie densities). Then let any two consecutive elements in any preference list from this set of agents be tied with probability t_d ($0 \leq t_d \leq 1$) (Kwanashie & Manlove, 2014). We reverse this procedure here, taking a real-world instance and calculating what proportion of ties exist on each side of it using the following process. First, count the number of distinct tie groups g and the number of actual elements e in the preference lists on one side, and let n be the number of agents on that side that have at least one agent in their preference list. The tie density of that side is then given by $t_d = 1 - \frac{g-n}{e-n}$. Subtracting n from the numerator ensures that if the agents consider all possibilities equally, we obtain a tie density of 1, and subtracting n from the denominator ensures that an absence of ties equates to a density of 0. Note that $e > n$ is assumed, that is, at least one agent has more than two agents in its preference list. If $e = n$, the instance is trivial to solve.

Through this formula, the tie density is 0.9716 for the families and 0.9705 for the children. Among the problem instances generated in Kwanashie and Manlove (2014), it was found that those with a tie density $t_d \approx 0.85$ tended to be the most challenging to solve.

For a given child or family, the variance of weights is relatively low: 3.8 on average for the children and 3.7 for the families. This suggests that some children are considered “good” or “bad” matches for many families, and vice-versa.

Each of the almost half a million child–family pairs is given a weight, but there are only 54 distinct weights in total. A bar chart displaying these weights is shown in Fig. 2 where we observe two distinct behaviours: some values appear many times (we will call them “common”) and some values appear just a few times in comparison (we will call them “uncommon”).

We examine our improvements to the models on the 22 SMTI-GRP real-world instances generated from the sample data set with the different thresholds. In all methods we considered, the preprocessing described in Section 4 was applied. Table 3 compares the six possible combinations of the proposed improvements. The “Method” columns detail the combination of options, with some attributes describing the specific implementation: “index” identifies the method while “dummy variables”, “stability constraint merging”, and “double stability constraints” indicate the inclusion or otherwise of the corresponding feature in the model. The three

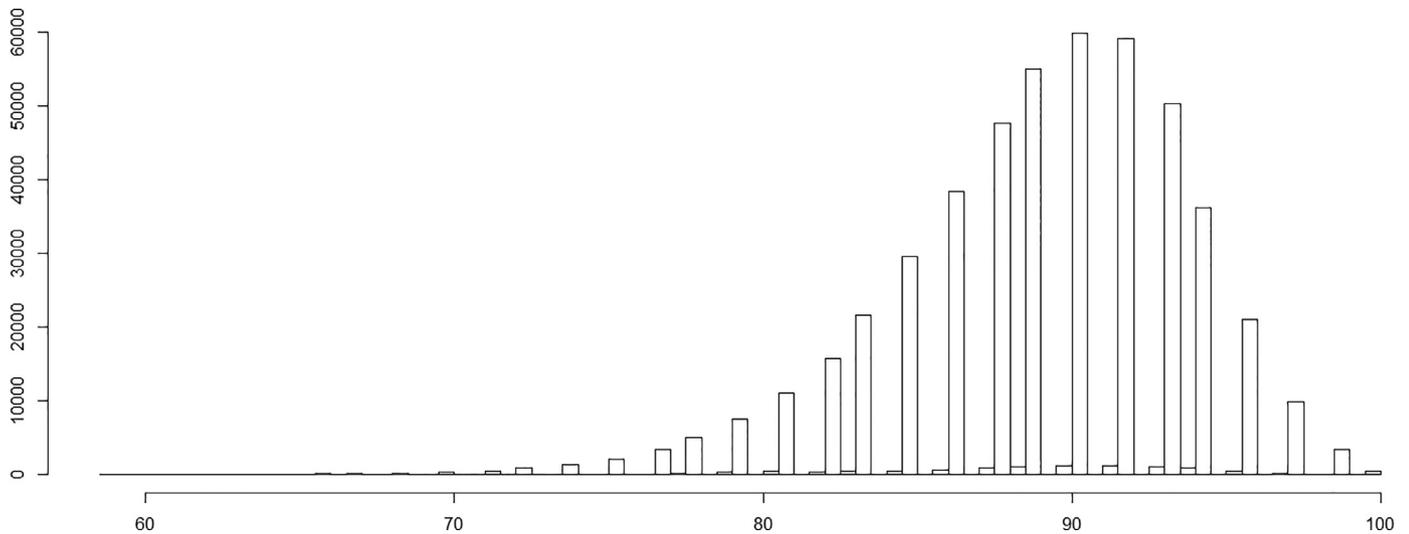


Fig. 2. Bar chart of the child-family weights for the complete instance.

Table 3

Comparison of the proposed methods for preprocessed SMTI-GRP real-world instances.

Method				Values			Model size		
Index	Dummy variables	Stab. cons. merging	Double stab. cons.	#Opt	Time	Continuous relaxation	Number of variables	Number of constraints	Number of non-zeros
M1				21	434.0	42966.0	94,764	96,208	39,152,977
M2	x			22	86.8	42966.0	101,220	101,220	390,790
M3		x		20	416.3	43030.4	94,764	3457	13,619,285
M4	x	x		22	73.7	43030.4	101,220	8468	298,038
M5		x	x	20	722.0	43010.0	94,764	7899	39,853,723
M6	x	x	x	22	66.2	43010.0	101,220	12,911	397,245

following columns give some indicators of the performance of each method: the number of optimal solutions found, average CPU time over all runs (including the ones terminated by the time limit), and the continuous relaxation value. The three last columns report some details about the model size: average number of variables, constraints, and non-zeros elements.

The table shows a number of interesting facts:

- The real-world instances did not challenge our algorithms as even the basic model M1, (introduced in Kwanashie & Manlove, 2014) solves 21 out of the 22 instances.
- Only the algorithms using dummy variables solve all the real-world instances. Algorithms using dummy variables are one order of magnitude faster than those that do not (e.g., 434 seconds on average for M1 versus 86.8 for M2). This is probably due to the decrease by two orders of magnitude in the average number of non-zero elements.
- Merging stability constraints seems to have a positive effect in terms of CPU time (434 seconds on average for M1 versus 416.3 for M3, and 86.8 seconds for M2 versus 73.7 for M4), but it remains marginal compared to the use of dummy variables. While using this feature worsens the continuous relaxation value (e.g., from 42966 for M1 to 43030.4 for M3), it also decreases the model size (e.g., from 96208 constraints for M1 to 3457 for M3).
- Provided that constraint merging is used, it is unclear if using double stability constraints is beneficial: while it increases the average time from 416.3 seconds for M3 to 722 for M5, it also decreases the average time from 73.7 seconds for M4 to 66.2 for M6. In contrast to constraint merging, the use of double stability constraints improves the continuous relaxation value at the expense of increasing the model size.

Supplementary experiments showed that, by dropping the stability constraints, the optimal objective value of our MAX-WT-SMTI-GRP real-world instances would increase by 3% on average (with a minimum of 0% and a maximum of 18%). Without stability constraints, the problem can be solved in polynomial time within seconds (Gabow & Tarjan, 1989).

Table 4 studies the impact of the preprocessing discussed in Section 4, the use of the Gale–Shapley algorithm (Gale & Shapley, 1962) to give a warm-start to the solver, and the use of priorities on the dummy variables during the branch-and-bound process. We tested these features on the basic model (M1), and one of the fastest algorithms (M4). We selected M4 because it obtained very good results while keeping a relatively low number of variables, constraints, and non-zero elements. In our experiments involving a warm start, we called the Gale–Shapley algorithm 100 times after breaking ties randomly, and fed the solver with the best solution found.

The results in Table 4 show that:

- Without preprocessing, the basic model M1 can only solve 6 out of 22 instances.
- Preprocessing leads to the removal of more than half of the variables, about half of the constraints, and more than half of the non-zero elements, both for models M1 and M4.
- Applying Gale–Shapley algorithm is useful as it reduces the average CPU time of M1 from 434 seconds to 288.9 and the average CPU time of M4 from 73.7 seconds to 59.2.
- Giving a high priority to the dummy variables y_{ik}^c and y_{jk}^f during the branching process of M4 does not help the solver. As this feature did not seem promising, we decided not to try further combinations of M4. We also mention that since M1 does not

Table 4
Comparison of M1 and M4 with additional features for SMTI-GRP real-world instances.

Method				Values		Model size		
Index	Preprocessing	Warm start	Dummy variables priorities	#Opt	Time	Number of variables	Number of constraints	Number of non-zeros
M1				6	1325.0	237,666	239,110	162,252,108
M1	x			21	434.0	94,764	96,208	39,152,977
M1	x	x		21	288.9	94,764	96,208	39,152,977
M4				22	155.4	248,442	15,689	738,240
M4	x			22	73.7	101,220	8468	298,038
M4	x	x		22	59.2	101,220	8468	298,038
M4	x		x	22	76.7	101,220	8468	298,038

Table 5
Comparison of the methods for preprocessed SMTI-GRP augmented instances.

Index	$\kappa = 1$						$\kappa = 2$					
	#Opt	Time	Cont. relax.	nb. var.	nb. cons.	nb. nzs.	#Opt	Time	Cont. relax.	nb. var.	nb. cons.	nb. nzs.
M1	220	340.6	45460.0	63,974	65,418	21,214,917	74	2701.5	92341.5	358,549	361,437	278,330,845
M2	220	114.8	45460.0	72,542	72,542	271,809	121	2185.1	92341.5	379,695	379,695	1,473,987
M3	220	201.5	45494.8	63,974	3770	6,742,990	52	2962.5	92476.6	358,549	8865	94,883,495
M4	220	84.0	45494.8	72,542	10,894	210,162	134	2114.1	92476.6	379,695	27,122	1,121,414
M5	220	507.9	45484.5	63,974	10,012	21,874,731	43	3046.8	92450.1	358,549	24,034	282,418,566
M6	220	55.9	45484.5	72,542	17,136	280,377	143	2106.7	92450.1	379,695	42,291	1,495,132

have dummy variables, the model could not be used to test this feature.

7.1.2. Augmented instances

As all real-world instances could be solved within an hour by the best method, we generated bigger instances to test the limits of our models. In order to keep the properties of the original instance, we used it as a basis to create the new benchmark.

We started by duplicating each child and each family κ times. Then, we perturbed the weight obtained by each pair in the following way: (1) identify the group of the weight (common or uncommon) and the position i of the weight in that group, and (2), change the weight to the one in position $i + x$ with probability $p(x)$ for $(x, p(x)) \in \{(-2, 0.1), (-1, 0.2), (0, 0.4), (1, 0.2), (2, 0.1)\}$.

For each $\kappa \in \{1, 2\}$, we generated 10 instances, to which we applied the 22 thresholds, resulting in 440 new instances in total. Table 5 compares the six possible combinations of the proposed improvements, with preprocessing, on the 440 augmented SMTI-GRP instances.

We observe that the comments made about the real-world instances are still valid here: using dummy variables reduces significantly the number of non-zero elements and improves the performance of the model. Constraint merging seems to have a positive effect, especially when used together with dummy variables. The same behaviour is observed for double stability constraints. Overall, we notice a clear improvement from the basic model M1 to the more sophisticated model M6, as the former could only solve 74 out of the 220 instances with $\kappa = 2$, while the latter could solve 143 of them. Despite this remarkable improvement, some instances with $\kappa = 2$ (i.e., with 1100 children and 1788 families) remain unsolved after one hour of computing time. We also generated instances with $\kappa = 3$, but running these experiments was impractical due to the large memory requirements of the models.

7.2. MAX-SMTI

Even though we initially developed our models for MAX-WT-SMTI-GRP, they can also be used to solve MAX-SMTI. As, to the best of our knowledge, no MAX-SMTI datasets are available in the literature, we used the generator described in Irving and Manlove (2009) to create new instances in order to test the effectiveness of

our methods on this problem too. We tried several combinations of number of agents ($\{10\,000, 25\,000, 50\,000\}$ on each side), tie density ($\{0.75, 0.85, 0.95\}$ on each side), and preference list length ($\{3, 5, 10\}$ on one side, as the generator does not support a limit on the preference list lengths on both sides). We generated 10 instances for each combination resulting in 270 SMTI instances in total.

Tables 6–8 compare the six possible combinations of the proposed improvements, with preprocessing, on the 270 random SMTI instances.

We observe that:

- As expected, for a given tie density and preference list length, a larger number of agents results in harder instances.
- For a given number of agents and a given preference list length, instances with a tie density at 0.75 are harder than those at 0.85, which are themselves harder than those at 0.95. This could be explained by the fact that the difference between the continuous relaxation and the optimal solution size is smaller as the tie density increases. For example, for M1 with 25,000 agents and a preference list length of 3, we observed an average absolute difference between the two values of 5.07 when the tie density was 0.75, 2.81 when the tie density was 0.85, and 0.15 when the tie density was 0.95.
- For a given number of agents and tie density, shorter preference lists make the instances easier. This is probably due to the fact that the models have fewer variables, constraints, and non-zero elements.
- Once again, the effectiveness of the new models is demonstrated: out of the 30 instances with preference list length of 5 and 50,000 agents, M1 solves 20 instances in total while M3 solves all 30 instances.
- For each combination, the best results were obtained either by M3 or by M4. We distinguish several cases:
 - When the preference list length is equal to 3, M3 is always the best. Indeed, the preference lists are too short to obtain any benefit from the dummy variables. For example:
 - For tie density 0.85, preference list length of 3, and 25,000 agents, M3 (resp. M4) has on average 72,227 variables (resp. 133,976), 81,322 constraints (resp. 93,071), and 408,556 non-zero elements (resp. 322,817, i.e., 21% less).

Table 6

Comparison of the methods for preprocessed SMTI instances with 10,000 agents.

Index	Preference list length = 3						Preference list length = 5						Preference list length = 10					
	0.75		0.85		0.95		0.75		0.85		0.95		0.75		0.85		0.95	
	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time
M1	10	10	10	8	10	8	10	155	10	94	10	39	0	3600	4	3428	10	382
M2	10	11	10	9	10	7	10	361	10	134	10	24	0	3600	0	3600	10	288
M3	10	5	10	5	10	5	10	75	10	38	10	23	9	2731	9	1555	10	490
M4	10	10	10	9	10	7	10	150	10	43	10	18	1	3543	10	1723	10	95
M5	10	10	10	8	10	8	10	115	10	74	10	41	0	3601	10	2073	10	525
M6	10	10	10	8	10	6	10	442	10	163	10	29	0	3600	0	3600	10	496

Table 7

Comparison of the methods for preprocessed SMTI instances with 25,000 agents.

Index	Preference list length = 3						Preference list length = 5						Preference list length = 10					
	0.75		0.85		0.95		0.75		0.85		0.95		0.75		0.85		0.95	
	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time
M1	10	65	10	53	10	45	10	1424	10	686	10	199	0	3600	0	3601	4	3439
M2	10	61	10	48	10	35	6	3329	10	928	10	144	0	3600	0	3600	6	3310
M3	10	23	10	24	10	23	10	583	10	230	10	117	0	3600	0	3601	3	3234
M4	10	56	10	46	10	32	10	1527	10	344	10	89	0	3600	0	3600	10	739
M5	10	55	10	46	10	38	10	873	10	429	10	170	0	3601	0	3601	7	2960
M6	10	54	10	44	10	31	0	3600	10	1449	10	156	0	3600	0	3600	0	3600

Table 8

Comparison of the methods for preprocessed SMTI instances with 50,000 agents.

Index	Preference list length = 3						Preference list length = 5						Preference list length = 10					
	0.75		0.85		0.95		0.75		0.85		0.95		0.75		0.85		0.95	
	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	Time
M1	10	230	10	202	10	146	0	3600	10	2565	10	523	0	3601	0	3601	0	3601
M2	10	239	10	209	10	146	0	3600	2	3336	10	407	0	3600	0	3600	0	3601
M3	10	74	10	78	10	73	10	1966	10	828	10	291	0	3601	0	3601	0	3602
M4	10	220	10	187	10	134	0	3600	10	1046	10	220	0	3600	0	3600	10	2132
M5	10	212	10	185	10	142	10	2684	10	1227	10	339	0	3597	0	3601	0	3603
M6	10	209	10	190	10	132	0	3600	0	3600	10	401	0	3600	0	3600	0	3600

- When the preference list length is equal to 5, M3 is always the best for tie densities 0.75 and 0.85 while M4 is always the best for tie density 0.95. This could be explained by the fact that high tie densities involve fewer tie groups. Thus, fewer additional constraints and variables are required by the dummy variables and using them saves more non-zero elements. For example:

- For tie density 0.85, preference list length of 5, and 25,000 agents, M3 (resp. M4) has on average 119,551 variables (resp. 196,273), 88 448 constraints (resp. 115,171), and 875,334 non-zero elements (resp. 500,736, i.e., 43% less).
- For tie density 0.95, preference list length of 5, and 25,000 agents, M3 (resp. M4) has on average 124,223 variables (resp. 183,693), 79,797 constraints (resp. 89,267), and 965,056 non-zero elements (resp. 471,574, i.e., 51% less).

- When the preference list length is equal to 10, M3 is the best for tie density 0.75, while M4 is the best for tie densities 0.85 and 0.95. Again, the use of dummy variables is shown to be beneficial with longer preference lists.

We also report that additional computational experiments showed that: (i) the preprocessing techniques introduced in Section 4 have little effect on these instances, and (ii) dropping the stability constraints would increase the size of the matching by at most 1% and make the problem polynomial-time solvable (Hopcroft & Karp, 1973).

7.3. MAX-HRT

In many instances of SMTI and HRT, it can be assumed that agents establish their ranking based on their own individual preferences. However, sometimes it is the case that agents' preferences are formulated on the basis of objective criteria. For example, in a specific version of MAX-HRT, hospitals only consider the grades of the doctors for their preference lists. In this situation, the so-called *master list*, a ranking of all the doctors based on their grades, is made at the beginning. The preference list of each hospital is then an exact copy of the master list from which the doctors who did not apply to the given hospital were removed. We tested our algorithms on instances both with and without a master list of doctors.

7.3.1. Non master list instances

We had access to instances of the Scottish Foundation Allocation Scheme, which assigned medical graduates to Scottish hospitals, for the years 2006, 2007, and 2008 (Irving & Manlove, 2009). These 3 instances, called "SFAS" in the following, have respectively 759, 781 and 748 doctors, 53, 53 and 52 hospitals, and 801, 789 and 752 available positions. Doctors chose exactly 6 hospitals, although a small number of exceptions with shorter preference lists were found. No master list was used by the hospitals to establish their preference lists, even if we observed a tendency for some doctors to be often well (or badly) ranked. The tie density on the hospitals' side was 0.9468, 0.7861 and 0.8424 respectively. The tie density on the doctors' side was 0 as the doctors were asked to provide strict preferences.

Table 9
Comparison of the proposed methods for preprocessed HRT SFAS instances.

Method				Values			Model size		
Index	Dummy variables	Stability constraints	Stab. cons. merging	#Opt	Time	Continuous relaxation	Number of variables	Number of constraints	Number of non-zeros
N1		(22)		3	144.8	747.4	1898	2714	63,272
N2	x	(22)		3	18.7	747.4	4146	4146	11,278
N3		(26)–(30)		3	10.1	744.5	2300	5014	16,495
N4	x	(26)–(30)		3	9.6	744.5	4548	6446	15,879
N5		(26)–(30)	(28) → (31)	3	15.9	747.8	2300	3465	14,946
N6	x	(26)–(30)	(28) → (31)	3	12.8	747.8	4548	4897	14,330
N7		(22) and (26)–(30)		3	16.1	744.3	2300	6912	75,971
N8	x	(22) and (26)–(30)		3	5.4	744.3	4548	8345	19,676
N9		(22) and (26)–(30)	(28) → (31)	3	25.5	746.2	2300	5363	74,422
N10	x	(22) and (26)–(30)	(28) → (31)	3	9.3	746.2	4548	6796	18,127
N11		(26)–(30) and (32)		3	5.6	744.3	2300	5311	21,377
N12	x	(26)–(30) and (32)		3	3.3	744.3	4548	6743	16,472

We also tested our algorithms on the dataset “SET2” described in Kwanashie and Manlove (2014) and available at <http://researchdata.gla.ac.uk/244/>. It comprises 700 instances with 100, 150, 200, 250, 300, 350, 400 doctors, and 7, 10, 14, 17, 21, 24, 28 hospitals, respectively. The number of available positions was exactly equal to the number of doctors. Doctors chose exactly 5 hospitals, and the tie density was equal to 0.85.

Table 9 compares twelve combinations of the proposed improvements, with preprocessing, on the 3 SFAS real-world instances. The meaning of the columns is unchanged with respect to Table 3, except for “stability constraints”, which now indicates the set of stability constraints used by the model, and “stability constraint merging”, which now indicates whether constraints (28) were merged or not.

The table shows a number of interesting facts:

- The real-world instances did not challenge the algorithms.
- Using dummy variables seems beneficial as algorithms that use them have always a lower average running time than those that do not (e.g., 144.8 seconds for N1 versus 18.7 for N2 and 5.6 seconds for N11 versus 3.3 for N12). Even if the decrease in the average number of non-zero elements is less spectacular than for SMTI-GRP, it is still significant for some models (e.g., 63,272 for N1 versus 11,278 for N2 and 21,277 for N11 versus 16,472 for N12).
- The kind of stability constraints used by the model appears to have a significant impact on the results: while N1 uses on average 144.8 seconds to solve the SFAS instances, N3 (which uses only the new set of stability constraints) requires merely 10.1 seconds. N7, that uses both sets of stability constraints, is a bit worse, even if it has the best continuous relaxation value: 744.3 versus 747.4 for N1 and 744.5 for N3. N11, that replaces the original set of stability constraints (22) by (32), obtains the best relaxation and one of the best average running times.
- Merging stability constraints (28) is not beneficial on these instances, as all algorithms that merge stability constraints have a worse average running compared to those that do not (e.g., 15.9 seconds on average for N5 versus 10.1 for N3, and 25.5 seconds for N9 versus 16.1 for N7). This can be explained by the fact that almost no gain is obtained in terms of number of constraints (e.g., 3465 for N5 versus 5014 for N3), but a significant loss is observed in terms of continuous relaxation value (e.g., 747.8 for N5 versus 744.5 for N3).
- Overall, this computational experiment suggests that the best configurations are N8, N11, and N12. All of them use two sets of stability constraints and no constraint merging.

Again, if stability constraints are dropped, the problem becomes solvable in polynomial time (Gabow, 1983) and the matching size is increased by at most 1%.

Table 10 compares the same twelve combinations on the literature instances SET2. Overall, the SET2 instances cannot be considered very challenging as all algorithms apart from N1 can solve them to optimality in less than 5 seconds on average. In addition, we notice that the algorithms’ behaviour does not change significantly: (i) dummy variables still seem useful, even if for some configurations (N4 and N12) no significant change is observed, (ii) the kind of stability constraints used still has a significant impact on the overall results, and (iii) using constraint merging still deteriorates the overall results. We note also that no major difference in the continuous relaxation is observed among the models. This is probably due to the fact that in 598 instances out of 700, all doctors could be assigned to a hospital, so the continuous relaxation value and the optimal solution were identical.

As all the real-world and literature instances could be solved within an hour, we generated new instances with the instance generator described in Irving and Manlove (2009). These instances have $759 \times i$ doctors, $53 \times i$ hospitals, and $775 \times i$ available positions, where $i \in \{1, 2, 3, 5, 10\}$ and are called “RDMi” in the following. Doctors chose between 5 and 6 hospitals, and the tie density on the hospitals’ side was equal to 0.85. For each i , 30 instances were created, resulting in 150 instances in total. These parameters were chosen to mimic the real-world SFAS instances at a larger scale.

Table 11 compares the twelve combinations on the RDMi instances. In order to be concise, we only report in the table the number of optimal solutions found and the average running time for each method.

Besides the observations made previously, which are still valid overall, we clearly notice that using two sets of stability constraints is significantly faster, especially for RDM2 and RDM3. In addition, we observe that large instances are extremely difficult, as only four of the RDM5 instances could be solved within an hour of computing time per instance. We also report that none of the tested algorithms could solve any of the RDM10 instances, even though the models had a reasonable size (e.g., for RDM10, N12 used on average 40,866 variables, 60,504 constraints and 145,259 non-zero elements). Finally, even if no “new” configuration clearly outperforms the others, all of them outperform the literature state-of-the-art model N1.

Table 12 studies the impact of the preprocessing developed in Irving and Manlove (2009) (we recall that the algorithms given in Section 4 are applicable to SMTI instance only), the use of the Gale–Shapley algorithm to give a warm-start to the solver, and the

Table 10
Comparison of the proposed methods for preprocessed HRT SET2 instances.

Method				Values			Model size		
Index	Dummy variables	Stability constraints	Stab. cons. merging	#Opt	Time	Continuous relaxation	Number of variables	Number of constraints	Number of non-zeros
N1		(22)		694	59.7	249.9	619	886	16,046
N2	x	(22)		700	1.8	249.9	1382	1382	3735
N3		(26)–(30)		700	0.7	249.9	781	1667	5315
N4	x	(26)–(30)		700	0.8	249.9	1544	2163	5297
N5		(26)–(30)	(28) → (31)	700	4.9	249.9	781	1193	4841
N6	x	(26)–(30)	(28) → (31)	700	1.6	249.9	1544	1689	4823
N7		(22) and (26)–(30)		700	1.3	249.9	781	2286	20,123
N8	x	(22) and (26)–(30)		700	0.6	249.9	1544	2782	6534
N9		(22) and (26)–(30)	(28) → (31)	700	3.1	249.9	781	1812	19,649
N10	x	(22) and (26)–(30)	(28) → (31)	700	0.8	249.9	1544	2308	6060
N11		(26)–(30) and (32)		700	0.7	249.9	781	1794	7706
N12	x	(26)–(30) and (32)		700	0.7	249.9	1544	2291	5552

Table 11
Comparison of the proposed methods for preprocessed HRT RDM instances.

Method				RDM1		RDM2		RDM3		RDM5	
Index	Dummy variables	Stability constraints	Stab. cons. merging	#Opt	Time	#Opt	Time	#Opt	Time	#Opt	time
N1		(22)		25	848	12	2656	2	3485	0	3600
N2	x	(22)		29	336	17	2056	9	2919	2	3524
N3		(26)–(30)		30	82	16	1844	9	2667	2	3491
N4	x	(26)–(30)		30	136	18	1815	10	2694	2	3395
N5		(26)–(30)	(28) → (31)	28	418	11	2358	4	3253	1	3574
N6	x	(26)–(30)	(28) → (31)	28	335	13	2253	6	3009	0	3600
N7		(22) and (26)–(30)		30	66	25	1170	14	2298	1	3495
N8	x	(22) and (26)–(30)		30	62	25	1070	13	2267	4	3372
N9		(22) and (26)–(30)	(28) → (31)	30	134	20	1702	13	2763	1	3509
N10	x	(22) and (26)–(30)	(28) → (31)	30	52	25	1023	14	2184	2	3397
N11		(26)–(30) and (32)		30	46	24	1270	13	2352	3	3434
N12	x	(26)–(30) and (32)		30	99	24	1252	14	2251	3	3353

Table 12
Comparison of N1 and N8 with additional features on RDM1 and RDM2.

Method				RDM1					RDM2				
Index	Prep.	Warm start	Variable priorities	#Opt	Time	nb. var.	nb. cons.	nb. nzs.	#Opt	Time	nb. var.	nb. cons.	nb. nzs.
N1				1	3581	4173	4985	225,332	0	3600	8343	9967	453,286
N1	x			25	848	1614	2426	37,563	12	2656	3260	4884	76,135
N1	x	x		28	524	1614	2426	37,563	15	2271	3260	4884	76,135
N8				30	42	9755	18,102	43,738	26	1036	19,470	36,156	87,373
N8	x			30	62	4075	7304	17,026	25	1070	8214	14,733	34,365
N8	x	x		30	82	4075	7304	17,026	25	957	8214	14,733	34,365
N8	x		z_{jk}	30	37	4075	7304	17,026	26	840	8214	14,733	34,365
N8	x		y_{ik}^d and y_{jk}^h	30	82	4075	7304	17,026	25	1069	8214	14,733	34,365

use of priorities on some variables during the branch-and-bound process. We tested these features on the basic model (N1), and one of the best algorithms (N8). We selected N8 because it obtained good results on all the datasets we tested in comparison with the other configurations (N8 solved 72 RDMi instances versus 71 for N12 and 70 for N11).

Unlike for MAX-WT-SMTI-GRP, the preprocessing seems useful for N1 but not for N8. Further investigations showed that, for N8, the inner preprocessing of Gurobi removed a similar amount of variables, constraints, and non-zero elements compared with the preprocessing of Irving and Manlove (2009). This was neither the case for N1, nor for M1 and M4 for MAX-WT-SMTI-GRP (with the preprocessing of Section 4). Using the Gale–Shapley algorithm to provide a warm start allowed N1 to solve an additional six instances from RDM1 and RDM2, but it slightly slowed down N8. Finally, giving priorities to the y_{ik}^d and y_{jk}^h variables during the branching process does not seem to help the ILP solver, however,

prioritising the z_{jk} variables appears to be beneficial. Further investigations showed that this statement was true for all the models involving z_{jk} variables (i.e., N3, N4, ..., N12).

7.3.2. Master list instances

As there is no existing set of instances that includes a master list in the literature, we used the generator described in Irving and Manlove (2009) to create new data sets. The same parameters ($759 \times i$ doctors, $53 \times i$ hospitals, and $775 \times i$ available positions, where $i \in \{1, 2, 3, 5, 10\}$) were used, and the grades obtained by the doctors were distributed in $[1, j]$ where $j \in \{5, 15, 25\}$. The distribution of doctor grades was controlled using a “skewedness” parameter x in the instance generator, which means that the most common doctor score is likely to occur x times more than the least common. Higher values of x hence result in longer ties in the master list, and therefore also in the hospitals’ preference lists. In our experiments we used the value $x = 3$.

Table 13
Comparison of N1 and N8 for preprocessed HRT RDM-ML- $i-j$ instances.

Instances		N1					N8 + priorities on z_{jk} variables				
j	i	#Opt	Time	nb. var.	nb. cons.	nb. nzs.	#Opt	Time	nb. var.	nb. cons.	nb. nzs.
25	1	30	0.0	834	1646	9869	30	0.1	2768	4436	9747
	2	30	0.1	1658	3282	19,462	30	0.1	5528	8843	19,420
	3	30	0.1	2464	4900	28,656	30	0.2	8217	13,145	28,850
	5	30	0.2	4097	8157	47,663	30	0.3	13,688	21,881	48,011
	10	30	0.5	8191	16,311	95,056	30	0.8	27,356	43,738	95,971
15	1	30	0.1	913	1725	12,230	30	0.1	2738	4565	10,159
	2	30	0.3	1804	3428	23,945	30	0.2	5443	9051	20,117
	3	30	0.5	2712	5148	35,939	30	0.4	8176	13,600	30,233
	5	30	3.8	4534	8594	60,305	30	1.4	13,647	22,715	50,516
	10	30	10.7	9000	17,120	118,923	30	9.5	27,164	45,164	100,373
5	1	29	227.0	1688	2500	49,510	30	1.1	3868	7244	17,052
	2	29	270.5	3336	4960	97,526	30	7.0	7657	14,329	33,709
	3	22	1326.5	5098	7534	151,203	30	271.3	11,665	21,861	51,488
	5	17	1929.3	8297	12,357	239,442	28	521.7	19,047	35,641	83,829
	10	6	3034.2	16,963	25,083	494,286	21	1734.3	38,807	72,733	171,301

The constructed instances are called “RDM-ML- $i-j$ ” in the following. For each pair (i, j) , 30 instances were created, resulting in 450 instances in total.

Table 13 compares the literature algorithm N1 and algorithm N8 when priorities are given to the z_{jk} variables (its best configuration). In both cases, preprocessing was applied.

It appears that instances that have a master list are significantly easier than those that do not, as each method can solve at least one instance of each group, even for RDM-ML-10- j . In addition, it seems that datasets allowing a larger range for the grades (RDM-ML- $i-25$ and RDM-ML- $i-15$), are easier as even the basic algorithm N1 can solve them all in seconds. Difficult master list instances have a very narrow range of grades (e.g., $j = 5$), and many doctors and hospitals (e.g., $i = 5$ or 10). On these instances, we can appreciate the benefits of the proposed improvements, as N1 can only solve 6 RDM-ML-10-5 instances, while N8 with priorities can solve 21 of them. Overall, for the 90 RDM-ML-10- j instances generated, N1 can solve 66, whereas N8 with priorities can solve 81, an increase of 23%.

7.4. Summary of the experiments

We empirically showed that, overall, we could solve significantly larger instances of MAX-WT-SMTI-GRP, MAX-SMTI, and MAX-HRT when compared to the existing methods. However, we observed that each problem had its own peculiarities.

For our practical case of MAX-WT-SMTI-GRP (the Coram application), characterised by a medium number of agents and very long preference lists, it is of paramount importance to reduce the number of non-zero elements and, thus, it is crucial to use dummy variables. Indeed, when preference lists are very long, stability constraints involve many variables and the size of the models increases quickly. To a lesser extent, it is beneficial to decrease the number of constraints without deteriorating the continuous relaxation too much. Thus, using constraint merging and double stability constraints is useful. In conclusion, configurations M4 and M6 are the most suitable for the problem.

For our MAX-SMTI instances, characterised by a very large number of agents and shorter preference lists, it is vital to reduce the number of constraints and, thus, it is advised to use constraint merging. Indeed, under these conditions, the models involve many stability constraints, which can be difficult to tackle by ILP solvers, even if they do not involve as many variables as they did for MAX-WT-SMTI-GRP. To a lesser extent, it is beneficial to decrease the number of non-zero elements, but only when it is not at the expense of a significant increase in terms of variables and constraints. Thus, using dummy variables is beneficial when the

tie density is high and when the preference list length is not too small. For this problem, configurations M3 and M4 are the most suitable.

For our MAX-HRT instances with no master list, characterised by a medium number of agents and short preference lists, it is important to help the solver reduce the gap between the lower and upper bounds and, thus, it is advised to use two sets of stability constraints. Indeed, we observed that the solver struggled to solve some instances even when the size of the model was reasonable. For this problem, configurations N8 and N12 are the most suitable.

The MAX-HRT instances with a master list that we tested, which aimed to mimic the real-world SFAS instances at a larger scale, did not present any sort of challenge when the grade range was reasonably large (i.e., at least 15 in our experiments). For instances with a very narrow grade range (i.e., 5 in our experiments), using N8 with preprocessing and the adequate priorities is advised.

Finally, we observed that preprocessing was useful for MAX-WT-SMTI-GRP, but not always useful for MAX-HRT (e.g., with configuration N8) or for MAX-SMTI. In addition, we saw that using a warm start could help some algorithms (e.g., N1 for MAX-HRT), and slow down some others (e.g., N8 for MAX-HRT). Finally, we also empirically observed that giving priorities to some specific sets of variables during the branching process of the ILP solver could be beneficial (e.g., giving priority to the z_{jk} variables of N8 for MAX-HRT).

8. Conclusion

We described two algorithms for preprocessing instances of SMTI where ties occur on both sides. This resulted in significant improvements when applied to models from the literature, solving an additional 15 (of 22 total) real-world MAX-WT-SMTI-GRP instances from Coram within one hour per instance. We also introduced new ILP models for SMTI, first by using dummy variables to reduce the complexity of the constraint matrix, and then by merging stability constraints, and using double stability constraints. Various combinations of techniques were demonstrated to improve the performance of our models, and together with the earlier preprocessing our new models solved all 22 Coram instances with a mean runtime of less than one minute. Computational experiments on randomly generated instances also showed that our models could solve instances of SMTI with up to 50,000 agents per side. The new ILP models were also extended to HRT, where we showed a performance improvement from 144 seconds to 3 seconds on average on real-world instances from SFAS. We also showed that we could solve an additional 23% of randomly gener-

ated instances with around 7500 doctors and hospital places when compared to state-of-the-art models.

In this paper we have not considered issues of strategy, however this direction is certainly worthy of further study in instances of SMTI-GRP and HRT. In the context of the classical Hospitals/Residents problem, it is a well-known result that, with respect to the Resident-oriented Gale–Shapley algorithm (Gale & Shapley, 1962; Gusfield & Irving, 1989), it is a dominant strategy for the doctors to tell the truth (Roth, 1985). On the other hand, there is no mechanism that is stable and strategy-proof for hospitals (Roth, 1986). Relative to any mechanism in SMTI-GRP or HRT that is based on finding a maximum cardinality stable matching, it is not difficult to show that weights or preferences (respectively) could be falsified by doctors and/or hospitals in order to improve their outcomes relative to their true preferences (e.g., by declaring less desirable preferences as unacceptable). We leave as future work the investigation of the existence of strategy-proof mechanisms for instances of SMTI-GRP and HRT that produce good approximations to maximum cardinality stable matchings.

Further work also includes extending our preprocessing algorithms for SMTI to the more general case of HRT. This may lead to the exact solution of larger HRT instances than we considered in this paper. It also remains open to extend our models to the extension of HRT where couples apply jointly to pairs of hospitals.

Acknowledgements

We would like to thank the two anonymous reviewers for their valuable comments that have helped to improve the presentation of this paper. We would also like to thank Kevin Yong of Coram for the collaboration involving the pairing of children with families, and also Rob Irving for helpful advice regarding the usage of the “skewedness” parameter that was used when generating HRT instances with master lists. This research was supported by the Engineering and Physical Science Research Council through grant nos. EP/P029825/1 (first four authors) and EP/P028306/1 (fifth and sixth authors).

References

- Abraham, D. J., Levavi, A., Manlove, D. F., & O'Malley, G. (2008). The stable roommates problem with globally-ranked pairs. *Internet Mathematics*, 5(4), 493–515.
- Ágoston, K. C., Biró, P., & McBride, I. (2016). Integer programming methods for special college admissions problems. *Journal of Combinatorial Optimization*, 32(4), 1371–1399.
- Arkin, E. M., Bae, S. W., Efrat, A., Okamoto, K., Mitchell, J. S. B., & Polishchuk, V. (2009). Geometric stable roommates. *Information Processing Letters*, 109, 219–224.
- Deligkas, A., Mertzios, G. B., & Spirakis, P. G. (2017). The computational complexity of weighted greedy matching. In *Proceedings of the thirty-first AAAI conference on artificial intelligence* (pp. 466–474).
- Delorme, M., Iori, M., & Martello, S. (2016). Bin packing and cutting stock problems: Mathematical models and exact algorithms. *European Journal of Operational Research*, 255, 1–20.
- Diebold, F., & Bichler, M. (2017). Matching with indifferences: A comparison of algorithms in the context of course allocation. *European Journal of Operational Research*, 260(1), 268–282.
- Drummond, J., Perrault, A., & Bacchus, F. (2015). SAT is an effective and complete method for solving stable matching problems with couples. In *Proceedings of the twenty-fourth IJCAI international joint conference on artificial intelligence* (pp. 518–525). AAAI Press.
- Erdil, A., & Erkin, H. (2008). What's the matter with tie-breaking? Improving efficiency in school choice. *American Economic Review*, 98, 669–689.
- Gabow, H. N. (1983). An efficient reduction technique for degree-constrained subgraph and bidirected network flow problems. In *Proceedings of the fifteenth STOC annual ACM symposium on theory of computing* (pp. 448–456). ACM.
- Gabow, H. N., & Tarjan, R. E. (1989). Faster scaling algorithms for network problems. *SIAM Journal on Computing*, 18, 1013–1036.
- Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, 69, 9–15.
- Gale, D., & Sotomayor, M. (1985). Some remarks on the stable matching problem. *Discrete Applied Mathematics*, 11, 223–232.
- García, S., Labbé, M., & Marín, A. (2011). Solving large p -median problems with a radius formulation. *INFORMS Journal on Computing*, 23, 546–556.
- Gent, I. P., Irving, R. W., Manlove, D. F., Prosser, P., & Smith, B. M. (2001). A constraint programming approach to the stable marriage problem. In *Proceedings of the seventh CP international conference on principles and practice of constraint programming*. In *Lecture Notes in Computer Science: 2239* (pp. 225–239). Springer.
- Gent, I. P., & Prosser, P. (2002). SAT encodings of the stable marriage problem with ties and incomplete lists. In *Proceedings of the fifth SAT international symposium on the theory and applications of satisfiability testing* (pp. 133–140). <https://ipg.host.cs.st-andrews.ac.uk/papers/gentprosserSAT2002.pdf> (accessed 1 August 2018).
- Gusfield, D., & Irving, R. W. (1989). *The stable marriage problem: Structure and algorithms*. MIT Press.
- Halldórsson, M., Iwama, K., Miyazaki, S., & Yanagisawa, H. (2007). Improved approximation results for the stable marriage problem. *ACM Transactions on Algorithms*, 3(3), 1–18. Article number 30.
- Hinder, O. (2015). The stable matching linear program and an approximate rural hospital theorem with couples. In *Proceedings of the eleventh WINE conference on web and internet economics*. In *Lecture Notes in Computer Science: 9470* (p. 433). Springer.
- Hopcroft, J. E., & Karp, R. M. (1973). A $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing*, 2, 225–231.
- Irving, R. W. (1994). Stable marriage and indifference. *Discrete Applied Mathematics*, 48, 261–272.
- Irving, R. W., & Manlove, D. F. (2009). Finding large stable matchings. *ACM Journal of Experimental Algorithmics*, 14, Section 1, 1–3. Article number 2.
- Irving, R. W., Manlove, D. F., & O'Malley, G. (2009). Stable marriage with ties and bounded length preference lists. *Journal of Discrete Algorithms*, 7(2), 213–219.
- Iwama, K., & Miyazaki, S. (2016). Stable marriage with ties and incomplete lists. In M.-Y. Kao (Ed.), *Encyclopedia of algorithms* (pp. 2071–2075). Springer. doi:10.1007/978-1-4939-2864-4_805.
- Király, Z. (2013). Linear time local approximation algorithm for maximum stable marriage. *Algorithms*, 6(3), 471–484. doi:10.3390/a6030471.
- Koné, O., Artigues, C., Lopez, P., & Mongeau, M. (2011). Event-based MILP models for resource-constrained project scheduling problems. *Computers & Operations Research*, 38, 3–13.
- Kwanashie, A. (2015). *Efficient algorithms for optimal matching problems under preferences*. University of Glasgow. Ph.D. thesis.
- Kwanashie, A., & Manlove, D. F. (2014). An integer programming approach to the Hospitals/Residents problem with Ties. In *Proceedings of or 2013 international conference on operations research* (pp. 263–269). Springer.
- Manlove, D. F. (2013). *Algorithmics of matching under preferences*. World Scientific.
- Manlove, D. F. (2015). Hospitals/Residents problem. In M.-Y. Kao (Ed.), *Encyclopedia of algorithms* (pp. 1–6). Springer.
- Manlove, D. F., Irving, R. W., Iwama, K., Miyazaki, S., & Morita, Y. (2002). Hard variants of stable marriage. *Theoretical Computer Science*, 276(1–2), 261–279.
- Manlove, D. F., McBride, I., & Trimble, J. (2017). “Almost-stable” matchings in the Hospitals/Residents problem with Couples. *Constraints*, 22(1), 50–72.
- Manlove, D. F., O'Malley, G., Prosser, P., & Unsworth, C. (2005). A constraint programming approach to the Hospitals/Residents problem. In *Proceedings of the fourth workshop on modelling and reformulating constraint satisfaction problems, held at CP '05: the eleventh international conference on principles and practice of constraint programming* (pp. 28–43). Springer.
- McBride, I. (2015). *Complexity results and integer programming models for Hospitals/Residents problem variants*. University of Glasgow. Ph.D. thesis.
- McDermid, E. (2009). A $3/2$ approximation algorithm for general stable marriage. In *Proceedings of the thirty-sixth ICALP international colloquium on automata, languages and programming*. In *Lecture Notes in Computer Science: 5555* (pp. 689–700). Springer.
- O'Malley, G. (2007). *Algorithmic aspects of stable matching problems*. University of Glasgow, Department of Computing Science. Ph.D. thesis.
- Paluch, K. (2014). Faster and simpler approximation of stable matchings. *Algorithms*, 7(2), 189–202. doi:10.3390/a7020189.
- Roth, A. E. (1985). The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36, 277–288.
- Roth, A. E. (1986). On the allocation of residents to rural hospitals: A general property of two-sided matching markets. *Econometrica*, 54, 425–427.
- Roth, A. E., Sönmez, T., & Ünver, M. U. (2004). Kidney exchange. *Quarterly Journal of Economics*, 119(2), 457–488.
- Roth, A. E., & Sotomayor, M. A. O. (1990). Two-sided matching: A study in game-theoretic modeling and analysis. *Econometric society monographs: 18*. Cambridge University Press.
- Rothblum, U. G. (1992). Characterization of stable matchings as extreme points of a polytope. *Mathematical Programming*, 54, 57–67.
- Slaugh, V. W., Akan, M., Kesten, O., & Ünver, M. U. (2016). The Pennsylvania Adoption Exchange improves its matching process. *INFORMS Journal on Applied Analytics*, 46(2), 133–153.
- UK Foundation Programme. 2018 recruitment stats and facts report. Available from <http://www.foundationprogramme.nhs.uk/document/2018-stats-and-facts-report> (accessed 28 Aug 2018).
- Vande Vate, J. V. (1989). Linear programming brings marital bliss. *Operations Research Letters*, 8(3), 147–153.