University of Glasgow | School of Computing Science

Maximum least-unstable matchings using Integer Programming

Duncan Adamson (2073411)

December 5, 2018

ABSTRACT

When we consider matching problems we often look to find matchings that are stable, however in many variants such as the hospital/residents problem such a matching may not include the maximum possible number of agents. While finding a maximum cardinality matching can be done in polynomial time, we often want to find one which is "almost stable", containing either as few blocking agents or pairs as possible, with the problem of finding a matching satisfying one of these conditions being NP-hard. In this work we will provide theoretical results for the these matchings, providing bounds on the numbers of both blocking agents or pairs, depending on what we are looking to minimise. We also present a new IP *model for finding such a matching, and results from experiments found using this model. Finally we will present a heuristic algorithm based on our theoretical results, as well as results from those algorithms.*

1. INTRODUCTION

Matching problems are a heavily studied area of computer science with many applications in logistics and economics. One of the first and simplest of these is the *Stable Marriage problem* (SM), in which we have a set of men and a set of women, where each man has a preference list containing all the women in order that he finds them preferable to be matched to, and each woman has a preference list of all of the men in order she finds them acceptable. In this problem we consider a matching, M, to be a set of man-woman pairs where each man and woman appears in no more than one pair. A *blocking pair* is some man-woman pair who are not matched to each other in M, but prefer each other to their current partners in M. We consider a matching that has no blocking pairs to be stable, Gale and Shapley [6] showed it was possible in an instance of SM to find a maximum cardinality matching that is stable in polynomial time.

Generalisations of this problem allow us to more accurately model real world problems. One example of this is the case where men and women may find some members of the other set unacceptable, we refer to this as the *Stable Marriage Problem with Incomplete lists* (SMI). It is worth noting that stable matchings from instances of this problem may not be a maximum cardinality matching, however all stable matchings will be of the same size, and any agent that is unmatched in one of these will be unmatched in all of them [9]. We can find, using Gale and Shapley's algorithm, a stable matching in polynomial time for the generalisation to SMI. It is known that a maximum cardinality matching in a bipartite graph can be found in polynomial time, however it is NP-hard to find a maximum cardinality matching that has the minimum number of blocking pairs [4].

We may further generalise this problem to include ties in agents preference lists, obtaining the *Stable Marriage Problem with Incomplete Lists and Ties* (SMTI). In this case we define a blocking pair in a weakly stable matching as one where we have some man-woman pair who are not matched to each other in the current matching M, and strictly prefer each other to their current partners in M, in this case requiring that they are not in a tie with each other. There are other notions of stability for this problem, however for this project we will only be considering weakly stable matchings, which we will refer to simply as stable when considering matchings with ties. Finding a maximum cardinality stable matching in SMTI (MAX-SMTI) was shown to be NP-hard by Iwama et al. [11], however we may find a arbitrary cardinality weakly stable matching by breaking the ties and using the Gale-Shapley algorithm [17].

One very important generalisation of SMI is the *Hospital/Residents problem* (HR), where the agents are hospitals and residents, with each resident being matched to a single hospital and each hospital to some number of residents up to its capacity. As with SMI, all stable matchings have the same size [7, 20], which we may find in polynomial time using Gale and Shapley's algorithm [6], however finding a maximum cardinality matching with the minimum number of blocking pairs is NP-hard. Again we may add ties to this generalisation, for the *Hospital/Residents problem with Ties* (HRT). In this case the problems of finding a maximum stable matching (MAX-HRT) and a maximum cardinality matching with the minimum number of blocking pairs (MAX-CARD MIN-BP HRT) are both NP-hard [4, 10, 14], with MAX-HRT being approximable within a factor of 2 [14, 11], while MAX-CARD MIN-BP HRT is not approximable within $n^{1-\epsilon}, \epsilon > 0$ unless $P = NP$.

The notion of such "Almost Stable" matchings was first introduced for the related Stable Roommates problem by Abraham et al. [1], an instance of which, unlike SM and HR, may not admit a stable matching. In this case we consider a matching to be "Almost Stable" if it contains the minimum number of blocking pairs. While this problem is NP-hard, finding if a matching with some constant number of of blocking pairs, K, exists can be done in polynomial time when K is a constant [1]. This was extend to SMI and HR by Manlove et al [4], who looked to find matchings of maximum cardinality, proving that this problem was NP-hard to solve. For an SMI instance, Manlove et al. showed that we can find if a matching of maximum cardinality with some set number of blocking pairs K, where K is a constant, can be done in polynomial time, as well as finding a MAX-CARD MIN-BP matching for both SMTI and HRT when the preference lists on one side were restricted to length at most 2. Some experimental results on the number of blocking pairs admitted by MAX-CARD MIN-BP SMTI matchings were presented by Mittal [19], who used an augmenting path based approach to find maximum cardinality matchings. Using this technique, they found a general increase in the cost of blocking pairs for matchings as the size of the matchings increases towards maximum cardinality.

For finding solutions to these problems, one commonly employed technique is Integer Programming (IP). There have been several prominent examples of using IP models for both NP-hard problems in general [5, 2, 8] and specifically matching problems [3, 13, 15, 16]. Relevant to us is the model by Kwanashie and Manlove [13] for MAX-HRT, who importantly found the sizes of stable matchings for the instances of the 2006, 2007 and 2008 Scottish Foundation Allocation Scheme. Many other IP models have focused on finding stable solutions to the *Hospital/Residents Problem with Couples* (HRC), a generalisation of HR where some residents are in couples with a joint preference list containing pairs of hospitals for both members of the couple, as well as further generalisations of HRC such as the case with ties (HRCT). While this problem is not what we are focusing on in this projects, looking at models from these [3, 15] has provided some ideas for our own. One highly relevant model based on this problem is by Manlove et al. [16] who looked to find "Almost Stable" matchings in HRC, though it is worth noting these were not maximum cardinality matchings.

In this work we will present our theoretical results on the structure of blocking pairs and agents in MAX-CARD MIN-BP HRT and MAX-CARD MIN-BA SMTI respectively, as well as bounds on the minimum number of blocking pairs and agents for such a matching respectively. We also present a new IP model for finding a matching of a given cardinality with the minimum number of blocking pairs, as well as a heuristic algorithm based on our theoretical results. We also present some empirical results found using the model we present here as well as for a modified version used to find a matching of a given cardinality with the minimum number of blocking agents.

Our paper will be organised as follows: in Section 2 we will present our theoretical results, in Section 3 we present the main new IP model for our paper, in Section 4 we present our heuristic algorithm based on the theoretical results, in Section 5 we present our experimental set up in addition to how we implemented and tested our tools, in Section 6 we discuss the results of our empirical study, finally we provide concluding remarks in Section 7.

2. THEORETICAL RESULT

We will first look at the theoretical results that we have derived for the problem. Our primary result concerns the structure of *above stable cardinality matchings*, in particular that the number of blocking pairs for a matching k greater than a stable matching will be strictly more than that of a matching of cardinality $k-1$ greater than a stable matching. We'll also some further results on the number of blocking pairs in matchings, in particular focusing on an upper bound on the number of blocking pairs. In these proofs we will assume that we are looking at instances of SMTI.

We will first define some notation that we will be using.

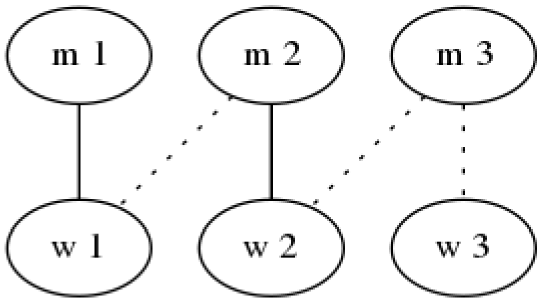

Figure 1: ***full lines indicated matched agents, dotted lines indicate blocking pairs***

We will define the problem of finding a matching of a given cardinality, K, with the minimum number of blocking pairs as K-CARD MIN-BP HRT and for a minimum number of blocking agents as K-CARD MIN-BA SMTI.

For denoting matchings, we will use $S(I)$ to denote a stable matching in an instance I of SMTI, and $\beta(I)$ to denote a maximum cardinality matching, with $\beta_{\text{BP}}(I)$ denoting a maximum cardinality matching with the minimum number of blocking pairs and $\beta_{\text{BA}}(I)$ denoting a maximum cardinality matching with the minimum number of blocking agents. For the number of blocking pairs in a matching, M, we will use $bp(M)$, and similarly for the number of blocking agents we use $ba(M)$. Finally, given a matching M and an agent A_i, we will us $M(A_i)$ to refer to the agent that A_i is matched to in M, if such an agent exists.

2.1 Blocking Pairs in K-CARD MIN-BP SMTI

In a matching of cardinality greater than stable we by necessity have blocking pairs between a man and a woman, which we will denote (m_i, w_j). There are three kinds of blocking pairs, which we will define here:

- ***Internal*** *(I)*: A blocking pair occurring between two agents within the matching
- ***Bridge*** *(B)* A blocking pair occurring between one agent within the matching and one unmatched agent
- ***External*** *(E)* A blocking pair occurring between two unmatched agents

We provide a visual example of the types of blocking pairs in Figure 1. In this case we have a matching containing the edges $(m1, w1)$ and $(m2, w2)$, with blocking pairs between $(m2, w1), (m3, w2)$ and $(m3, w3)$. In this case the blocking pair $(m2, w1)$ forms an internal blocking pair, $(m3, w2)$ forms a bridge blocking pair and $(m3, w3)$ forms an external blocking pair.

We will define *resolving* a blocking pair, (m_i, w_j), to be the process of removing any pairs, $(m_i, M(m_i)), (M(w_j), w_j)$, from the matching if they exist, then adding (m_i, w_j) to the matching. We may note here that when resolving an Internal blocking pair we reduce the cardinality by 1, resolving a bridge blocking pair maintains the cardinality, and an External blocking pair will increase the cardinality 1. The *best blocking pair* for some blocking agent is the blocking pair made with the agent it finds most preferable (or an agent it finds most preferable in the case of ties). We may also consider the best blocking pair conditional on it being a certain kind, most importantly bridge, in this case we consider it to be the most preferable blocking pair of that kind some agent makes. When resolving a blocking pair we will always consider there to be an initiating agent, which will be matched with the best blocking pair within the constraints of the type of blocking pair we are resolving.

For all blocking pairs, we will consider a structure within this matching to be a set consisting of some initial blocking agents, $A_1 \cdots A_k$, making either an Internal or bridge blocking pair, that when resolved will create some number of blocking pairs from $M(A_i), i \in 1 \cdots k$. By continuing to resolve these either we will end with some pair of agents unmatched who form no bridge blocking pairs, or will not terminate by continuing to resolve in this way, ending with some pair of agents we have already encountered forming bridge blocking pairs with other previously encountered agents. We will consider the first case to be a *chain* and the second case to be a *cycle*. The structure will consist of the set of the agents that are in the matching after we get by resolving any bridge blocking pairs from the unmatched agents.

One important aspect to a cycle is that if it includes a smaller cycle within it, we can always decompose this into a smaller cycle and a chain. We can see this as follows, assume we have some cycle containing within itself a cycle. In this case there will be some subset of the agents that will not be reached by resolving the unmatched agents from the cycle, giving us a chain and a smaller cycle. As such, we will only consider cycles that may not be decomposed in the remainder of this work, considering any chains from decompositions as entirely separate structures.

In the case of a chain we will only have a single initial agent, A_1, while in the case of a cycle we may have multiple initial agents $A_1 \cdots A_k$. We will consider an *I-Chain* to be a chain initiated by an Internal blocking pair, and an *I-Cycle* to be a cycle initiated by one ore more Internal blocking pairs. Similarly we have a *B-Chain* and *B-Cycle* as a chain or cycle initialised by a bridge blocking pair(s). We will first make two statements on these structures:

Lemma 2.1.1. *Any B-Chain may always be resolved without reducing the cardinality or adding to the number of internal blocking pairs.*

Proof. We can see this as any agents involved in the chain that remain in the matching will only have their partner improved as a result of the chain. As such given a matching of greater than stable cardinality, we can always find a matching of the same cardinality without any B-Chains. □

Lemma 2.1.2. *Any cycle must have a stable matching containing all the agents within it.*

Proof. We can see this by considering some cycle that admits no stable matching, in order for there to not be a stable matching there must be some pair of agents that we can remove without adding blocking pairs, contradicting that this would be a cycle as it would terminate. □

From this we can see that any I-Cycle can be resolved with the same cardinality and less blocking pairs, meaning that in a minimum blocking matching we can not have any I-Cycles.

Lemma 2.1.3. *Given any cycle, we can always create a matching of any cardinality up to the size of the cycle that will contain no internal blocking pairs.*

Proof. We can see this by considering the stable matching of maximum cardinality, from this a matching of smaller cardinality that is internally stable may be trivially created by removing any pair from this matching. □

Given these results, we will now claim the following:

Theorem 2.1.1. *In a* K-CARD MIN-BP SMTI *matching for some instance I, $K > |S(I)|$, all blocking pairs must be internal and may only be involved in I-Chains*

Proof. We first consider some matching of above stable cardinality with the minimum number of blocking pairs for that cardinality that consists of some number of I-Chains, bridge blocking pairs and External blocking pairs. We will first resolve any B-Chains in the matching, removing them from the matching. Next we may resolve some I-Chain, and any generated B-Chains, decreasing the cardinality of the matching by 1 and as any change we make to the matching will only increase the preference of agents with their matches we will in doing so decrease the number of internal blocking pairs. From this we may increase the cardinality of the matching by resolving adding the best blocking pair of some agent in a B-Cycle, which we may use Lemma 2.1.3 to ensures that the matching is internally stable. We may repeat this process until either we have no B-Cycles or no I-Chains.

In the case we have no B-Cycles, we have two disjoint sets of agents, in the case we still have external blocking pairs left we can continue to resolve some I-chain then add some external blocking pair to the matching, again resolving any B-Chains that may arise from it. We can repeat this until either we have no external blocking pairs or no I-Chains. In the case we have no external blocking pairs, as we have been strictly decreasing the number of internal blocking pairs, we must have less blocking pairs than in the initial matching, contradicting our assumption that it had the minimum number of blocking pairs for a matching of its cardinality. In the case that we have no I-chains, we may take advantage of the fact that this matching consists of two disjoint sets, finding a stable matching within the unmatched agents giving us a stable matching of greater cardinality than our initial matching and thus contradicting our assumption of the matching being above stable cardinality.

In the case we have some B-Cycles but no I-Chains, we may again add the B-Cycles into the matching, resolving any B-Chains that arise, until we again have either two disjoint sets of agents. As the matching will contain no I-Chains, by finding a stable matching within the unmatched agents, we again find a matching of greater cardinality that is stable, contradicting our original assumption. □

Given this, we can see that any matching of greater than stable cardinality with a minimum number of blocking pairs can only contain I-Chains. From this we can derive the following Theorem:

Theorem 2.1.2. *Given an instance, I, and two matchings of I, M and M', $|M'| > |M| > |S(I)|$, both with the minimum number of blocking pairs for a matching of their cardinality, $bp(M') > bp(M)$.*

PROOF. We can see this by considering the matching M'. As this matching must have some blocking pair, which must be internal, we may resolve it to decrease the cardinality of the matching by 1. If resolving this would generate a B-Cycle or a disjoint set of unmatched agents, we could find a matching of greater cardinality that is stable, therefore it may only generate B-Chains, which we may resolve to find a matching of cardinality, $M'', bp(M') > bp(M'')$. We may repeat this process $|M'| - |M|$ times, getting a matching of cardinality $|M|$ with strictly less blocking pairs than $|M'|$, though not strictly the same number as in M. $\square$

As such, given a matching M for some instance I, $|M| > S(I)$, we have that $bp(M) \geq |M| - |S(I)|$.

2.2 Blocking Agents in K-CARD MIN-BA SMTI

We will now claim that given an instance I and two matchings for I, M and M' $|M'| > |M| > S(I)$, such that both M and M' have the minimum number of blocking agents for their cardinality, $ba(M') > ba(M)$. In this case we will consider agents to be either *internal* or *external*, referring to whether the agent has a partner or not in the matching.

Lemma 2.2.1. *In a K-CARD MIN-BA matching there are no external blocking agents.*

PROOF. We can see this by considering some K-CARD MIN-BP SMTI matching, containing at least one external blocking agent. In this case we will have at least one bridge or external blocking pair. In the case we have bridge blocking pairs we may resolve them without adding any new internal blocking pairs as noted in Lemma 2.1.1. Again we can continue to resolve these until we have less internal blocking agents than previously and some number of external blocking agents, containing only external blocking pairs. We can resolve one of the external blocking pairs and internal blocking pairs simultaneously, maintaining the same cardinality of the matching, and decreasing the number of bridge blocking pairs which we can resolve. By repeating this process we will either reach the point where we have no internal blocking agents, allowing us to find a matching of greater cardinality that is stable, contradicting our assumption that the matching was of above stable cardinality, or a matching containing strictly less blocking agents all of which are internal, contradicting the assumption that the matching had the minimum number of blocking pairs. $\square$

Using this we claim:

Theorem 2.2.1. *For some instance, I, given two matchings, M and M', $|M'| > |M| > S(I)$, where both M and M' have the minimum number of blocking agents for a matching of their cardinality, $ba(M') > ba(M)$.*

PROOF. Given M', we may find some matching of cardinality $|M'| - 1$ by resolving it with its most preferable blocking pair, reducing the number of internal blocking pairs in the matching. We can subsequently resolve any bridge blocking pairs leaving us with only internal and external blocking pairs. In the case we have external blocking pairs, we may resolve one internal blocking pair and one external blocking pair simultaneously, until either we have no external blocking pairs and strictly fewer internal blocking agents than in M', or only external blocking pairs, allowing us to find a stable matching of greater cardinality contradicting our original assumption. By repeating this we reduce the number of blocking agents at each stage, proving that $ba(M') > ba(M)$. $\square$

Given some instance I, any matching $M, |M| > |S(I)|, ba(M) \geq |M| - |S(I)|$. We may find a stricter bound considering Theorem 2.2.1 alongside the following facts

Lemma 2.2.2. *There must be at least one agent in the matching that does not form a blocking pair in any matching with a minimum number agents*

PROOF. We can see this by considering the case where we have a matching where all agents form at least one blocking pair, we would have an alternating cycle consisting of an internal blocking pair edge followed by a matched edge. This would allow us to find a matching of that cardinality with less blocking pairs and agents, hence there must be at least one agent that is not involved in a blocking pair. $\square$

This provides an upper bound on the number of blocking agents in any K-CARD MIN-BA matching of $2k - 1$. We can also see this same limit applies for a K-CARD MIN-BP matching. We may further generalise this to

Lemma 2.2.3. *In a matching with the minimum number of blocking agents, there can be no alternating cycles between blocking edges and matched edges.*

We can trivially see this holds from the previous proof. From these we claim:

Theorem 2.2.2. *Given a matching M for some instance I, $|M| > S(I)$, with the minimum number of blocking pairs for its cardinality, there exists some matching of size $|M|-1$ containing 2 fewer blocking agents.*

PROOF. We know from Lemma 2.2.3 that all blocking agents may only be involved in chains, allowing us to select some blocking agent that is at one end of a chain, by following the alternating path from this agent along the most preferable edge we will eventually reach some point were two agents will share a most preferable edge. By resolving this, both agents will no longer be involved in blocking pairs, thus removing 2 agents from the matching. We know there can be no external blocking pairs as if there were we could find a matching of size $|M|$ with less blocking agents, invalidating our initial assumption. Any bridge blocking pairs must contain the two agents that were originally in the matching, allowing us to resolve them until we are left with none and at least 2 fewer blocking agents than in our original matching. $\square$

From this we get that given some instance I, any matching $M, |M| > |S(I)|$, $ba(M) \geq 2(|M| - |S(I)|)$

3. IP MODEL

We will now present our new IP model for finding a matching of some given cardinality K in an instance of SMTI. As input we will assume we have our instance, consisting of a set of men, U, and women W. The rank of some agent A_i on a second agent A_j's preference list is either the position of A_i on A_j's preference list, or the size of the set to which A_i belongs + 1. In the case we have two agents that are tied on some preference list, they share the same rank for that preference list. We will define $R(m_i, w_j)$ to return the rank of the j^{th} woman on the preference list of the i^{th} man, and similarly $R(w_j, m_i)$ which returns the rank of the i^{th} man on the j^{th} woman's preference list. We will also add a dummy man and dummy woman to the end of the sets of men and women respectively, we use these to represent unmatched agents.

Using these we will construct two *Rank Subtract Preference Matrices*, P^U and P^W which we define as follows:

$$P^U_{i,j} = (|W| + 1) - R(m_i, w_j) 1 \leq i \leq |U|, 1 \leq j \leq |W|$$
$$P^W_{i,j} = (|W| + 1) - R(w_j, m_i) 1 \leq i \leq |U|, 1 \leq j \leq |W|$$

Giving a value of 0 for unacceptable partners, or a positive value for acceptable partners such that $P^U_{i,j} > P^U_{i,p}$ if and only if the $i^t h$ man prefers the j^{th} woman to the p^{th} one, and similarly $P^W_{i,j} > P^W_{q,j}$ if and only if the j^{th} woman prefers the i^{th} man to the q^{th} one. We will also define two matrices of variables, the first will be M, where

$$M_{i,j} = \begin{cases} 1, & \text{if } m_i \text{ is matched to } w_j \\ 0, & \text{otherwise} \end{cases}$$

$$1 \leq i \leq |U| - 1, 1 \leq j \leq |W| - 1$$

And to represent unmatched agents

$$M_{i,|W|} = \begin{cases} 1, & \text{if } m_i \text{ is unmatched} \\ 0, & \text{otherwise} \end{cases}$$

$$1 \leq i \leq |U| - 1$$

$$M_{|U|,i} = \begin{cases} 1, & \text{if } w_j \text{ is unmatched} \\ 0, & \text{otherwise} \end{cases}$$

$$1 \leq j \leq |W| - 1$$

and a second matrix, B where

$$B_{i,j} = \begin{cases} 1, & \text{if } (m_i, w_j) \text{ form a blocking pair} \\ 0, & \text{otherwise} \end{cases}$$

$$1 \leq i \leq |U| - 1, 1 \leq j \leq |W| - 1$$

Given these variables, we subject them to the following constraints:

$$B_{i,j} \in \{0, 1\} \qquad 1 \leq i \leq |U| - 1, 1 \leq j \leq |W| - 1 \tag{1}$$

$$M_{i,j} \in \{0, 1\} \qquad 1 \leq i \leq |U| 1 \leq j \leq |W| \tag{2}$$

$$\sum_{j=1}^{|W|} M_{i,j} = 1 \qquad 1 \leq i \leq |U| - 1 \tag{3}$$

$$\sum_{i=1}^{|U|} M_{i,j} = 1 \qquad 1 \leq j \leq |W| - 1 \tag{4}$$

$$M_{i,j} \leq P^U_{i,j} \cdot P^W_{j,i} \qquad 1 \leq i \leq |U| - 1, 1 \leq j \leq |W| - 1 \tag{5}$$

$$\sum_{i=1}^{|U|-1} \sum_{j=1}^{|W|-1} m_{i,j} \geq K \tag{6}$$

$$1 \leq i \leq |U| - 1, 1 \leq j \leq |W| - 1, 1 \leq q \leq |W|, 1 \leq p \leq |U|$$
$$B_{i,j} \geq \frac{(P^U_{i,j} - P^U_{i,q}) M_{i,q}}{|(P^U_{i,j} - P^U_{i,q})| + 1} + \frac{(P^W_{i,j} - P^W_{p,j}) M_{p,j}}{|(P^W_{i,j} - P^W_{p,j})| + 1} - \frac{|U| + |W|}{|U| + |W| + 1} \tag{7}$$

With the objective to minimise the value of $\sum_{i=1}^{|U|-1} \sum_{j=1}^{|W|} B_{i,}$. In this we are assuming that the sets contain the dummy agents as the final member, and that all agents find them to be unacceptable. Constraints 1 and 2 ensure that both the matching and blocking pair matrices are binary valued. Constraints 3 and 4 ensure that each of the men and women, excluding the dummy cases, are matched to either to a member of the other set, or indicate they are unmatched. Constraint 5 ensure that agents are only matched if they each find the other acceptable, or are the dummy agent. Constraint 6 ensures that our matching is of cardinality at least K, which, as we are minimising the number of blocking pairs in the matching is equivalent to requiring the matching to be of size K as we know from Theorem 2.1.2.

Constraint 7 requires more intuition to understand. In this constraint we are checking, for each possible man-woman pair all other men and women to see if they are a blocking pair. The first fraction obtains us a positive value if and only if m_i is matched to w_q but finds w_j preferable. Trivially we can see that if he is not matched to w_q the fraction will return 0, allowing us to assume they are matched for the remainder of this discussion. In the case w_j is less preferable than w_q, the overall value of the fraction will be negative, being at most $\frac{-1}{2}$. In the case where the two are equally preferable the value will be 0, and in the case where w_j is more preferable the fraction will be at least $\frac{1}{2}$ and at most $\frac{|W|}{|W|+1}$. We can see the same arguments apply to the second fraction, with the upper bound on a positive value being $\frac{|U|}{|U|+1}$. As such the sum of these two parts will be greater than or equal to 1 in the case the two prefer each other to their currently matched partner in the matching, and at most the larger of $\frac{|W|}{|W|+1}$ and $\frac{|U|}{|U|+1}$ in the other case, either of which will be less than $\frac{|U|+|W|}{|U|+|W|+1}$, which is less than 1, giving us an overall sum of some value greater than 0 if there is a blocking pair, and less than 0 if not. The upper bound in the case where m_i and w_j is $\frac{|U|}{|U|+1} + \frac{|W|}{|W|+1} - \frac{|U|+|W|}{|U|+|W|+1}$, which we can see will be less than 1 in all cases.

We provide our extension of this model to further problems in Appendices A, B and C. For now we will simply note that these extend the model the MAX-CARD MIN-BP HRT, MAX-SMTI, MAX-HRT, as well as the cases for these problems we wish to minimise the number of blocking agents rather than blocking pairs.

4. HEURISTIC ALGORITHM

Moving beyond our IP model, we also provided a heuristic algorithm (Algorithm 1) based on our theoretical results. The first step is to find an arbitrary maximum cardinality matching. We do this using a simple augmenting path based approach. Once we have this matching, our next step is to find and resolve some bridge chain. We choose this path using a simple greedy approach where for each unmatched agent we check if it is involved in a bridge blocking pair, if it is we create a new path starting with this blocking pair followed by the matched edge adjacent to the matched agent in the initial blocking pair. From this we take the agent not adjacent to a blocking edge in the path, then add the best blocking edge containing this agent to the matching, and the adjacent matched edge. We repeat this until we reach an agent with no blocking edges, at which point we compare this to the best path found so far and, if found better by our weighting method, we replace with the new one. Having checked all unmatched agents, we resolve the best path found and repeat with the new matching. Once we have removed all bridge blocking pairs from the matching we search for any alternating cycles in the matching, resolving them when we find them.

For determining the best path, we devised 4 methods for weighting the paths:

- The length of the path
- The sum of the improvements for all agents in the path
- The sum of the improvement of the agent in each edge improving by the most
- The sum of the improvement of the agent in each edge improving by the least

In the first case we simply looked at the number of edges in our alternating path, selecting the path that was the longest as the best. In the second case we took the weight to be the sum of the difference in rank for the two agents in the blocking edge, and their current partner (or length of the preference list, in the case they were unmatched). For the third case we only took the value of the agent improving by the greatest amount, and in the fourth case by the least amount.

5. EXPERIMENTAL SET-UP AND PROCEDURE

In this section we will discuss the set up of the instances used for our experiments, as well as the tools used to obtain the results. We will also include a discussion of our experimental methodology and the verification testing for our tools.

```
Data: Graph G
Result: matching in G
find arbitrary maximum matching of G;
while G contains a bridge blocking pair do
    find best path starting with bridge blocking pair;
    resolve path;
end
Cycle p = find cycle in reduced graph;
while p ≠ null do
    resolve p;
    p = find cycle in reduced graph;
end
```

Algorithm 1: Heuristic Algorithm

5.1 Instances

We will first discuss the instances we used for our experiments, starting with what aspects of the instances we vary in our experiments, followed by our set up for instances of SMTI and HRT. For both SMTI and HRT instances we vary the following four aspects:

- *Instance Size:* The number of agents in the instance
- *Tie Density:* The probability that, the i^{th} agent in a preference list is tied with the $i+1^{th}$
- *Length of Preference Lists:* The upper bound on the length of preference lists for the agents
- *Skewedness:* The amount of preference lists the most popular agents occur in, as a ratio compared to that of the least popular ones

For SMTI instances our standard set up had 50 men and 50 women, with a tie density of 0 (making all instances where we were not varying the ties SMI instances), an upper bound of 3 on preference lists for both sides and a skewedness value of 5.

For our experiments, when varying instances size we increased the size for both the men and women in increments of 5. For tie densities we varied the tie density for the men from 0 to 1 in increments of 0.1, varying at each of these increments the density for the woman from the current density of the men to 1 again in increments of 0.1. We can safely ignore tie densities for women lower than that of the men due to the symmetries of the problem, meaning that the results for the men with tie density of i, and women with tie density of $j, j \geq i$ should be approximately the same given a sufficient sample size as instances with men of tie density of j and women with a tie density of i. For preference lists we vary the length from 1 to 5 in increments of 1 for the men, and varying at each increment, where the men have preference lists of length i, the preference lists of women from i to 5, again ignoring the symmetry for the same reasons as in the case of tie densities. Finally in the case of skewedness we vary the degree lists are skewed by from 2 to 10, varying in each case the skewedness of the men by 1, and for each of these the skewedness of the women from that of the men to 10.

For HRT instances our standard instances have 50 residents and 10 hospitals, with 60 posts randomly distributed between the hospitals. Again we have a tie density of 0 (making any instance where we're not varying the tie density an HR instance), an upper bound of 3 for the length of residents preference lists and unbounded for hospitals and a skewedness value of 5 for both sides.

For our experiments, when varying the size of instances we maintained a ratio of 5 residents for each hospital, and 1.1 posts for each resident randomly distributed. We varied the size of instance from 15 to 100 residents in increments of 5. For tie density we varied the density separately for hospitals and residents by 0.1 from 0 to 1, considering all combinations of hospital and resident tie densities. For the upper bound on the length of preference lists we varied the value for both hospitals and residents from 1 to 5, again considering all possible of combinations. Finally for the skewedness we independently varied the value for both hospitals and residents for 2 to 10 in increments of 1.

We also preformed experiments on scalability of our tools. In this case we used HR instances set up as before varying the size from 5 to 125, at which point our tools could no longer scale to find optimal solutions fast enough to solve a sufficient amount to draw conclusions. For verification testing we used instances of SMI of size 7 to solve with a brute force tool. The limit of 7 was chosen as any higher could not be efficiently solved using the brute force tool.

In our experiments we looked at finding solutions to find maximum stable, MAX-CARD MIN-BP, MAX-CARD MIN-BA, K-CARD MIN-BP and K-CARD MIN-BA matchings, where k was varied from one greater than the size of a stable matching to one less than the size of a maximum cardinality matching. In these cases we measured the cardinality of the matching, number of blocking pairs and the number of blocking agents. We also recorded the matching to allow verification to be done. Finally we also recorded the time, initially in seconds though this was improved to milliseconds when we were testing the scalability of the model.

5.2 Tools

For solving instances optimally we used our IP models presented in Section 3 and in Appendices A, B and C. We implemented these using Gurobi with the Java API. Gurobi was chosen due to the results from surveys on IP solvers [12, 18], suggesting that it was better than open source solvers and at least comparable to other commercial solvers. Java was chosen as the language to implement our tools in due to familiarity with the language. Our tool is operated from the command line, taking in the models, output file and instances as arguments as well as optional flags to specify how the tool is to be run. For solving the problem of finding a stable matching in SMI and HR instances, we use Gale and Shapley's algorithm, and an augmenting path algorithm to find a maximum cardinality matching.

We also implemented our heuristic algorithm, presented in Section 4, as a command line based Java tool. For the verification testing of our primary tool we created a back tracker in Java to check all possible solutions to the the instance.

To check our primary tool we generated 10,000 instances of SMI of size 7 with preference lists of size at most 3, and a skewedness of 5. The size of 7 was chosen as it was the largest size our brute force tool was able to solve quickly. We solved these instances for SMI, MAX-CARD MIN-BP SMI, MAX-CARD MIN-BA SMI, K-CARD MIN-BP SMI and K-CARD MIN-BA SMI with both the IP based solver and our brute force tool. In all cases our values from both tools were equal. For additional verification we also checked against the instances previously solved by Kwanashie and Manlove [13] for MAX-HRT for the instances of size up to 200 for our IP solver.

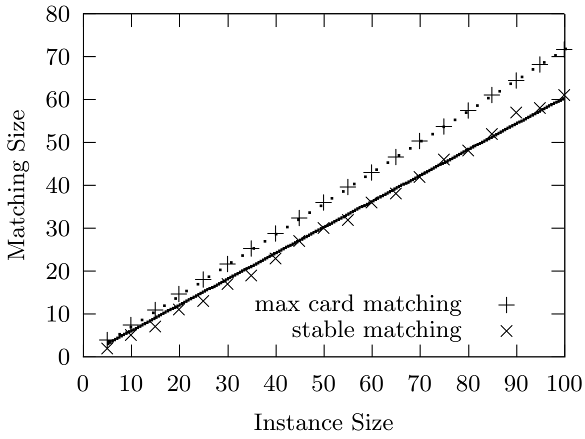

Figure 2: ***Instance Size versus Average Matching Size for stable and maximum cardinality matchings***

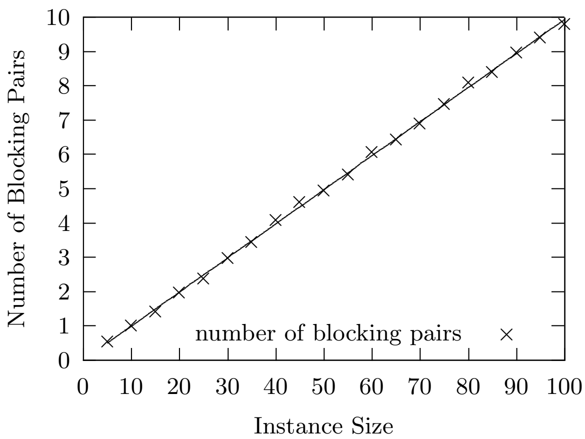

Figure 3: ***Instance Size versus Average Number of Blocking Pairs in a*** MAX-CARD MIN-BP SMTI ***matching***

6. EXPERIMENTAL RESULTS

In this section we will look at our results for solving the instances. We will first look at the optimal solutions to instances, followed by our tests on real instances, our heuristic results, and finally our scalability results.

6.1 Optimal solutions for SMTI

We will look at our experiments for SMTI in the following order, size of matchings, tie density, preference list length and finally skewedness, where the instances were set up as described in section 5. In all cases we used 1000 randomly generated instances so as to get a reasonable sample for us to draw conclusions.

In Figures 2, 3 and 4 we show our results for the cardinality of a stable and maximum cardinality matchings, the number of blocking pairs in MAX-CARD MIN-BP SMTI match-

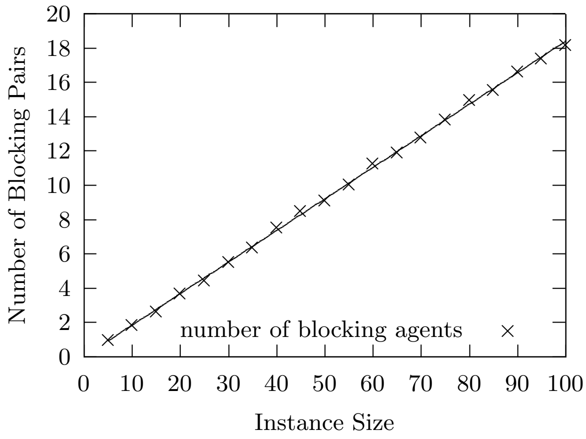

Figure 4: ***Instance Size versus Average Number of Blocking Agents in a*** MAX-CARD MIN-BA SMTI ***matching***

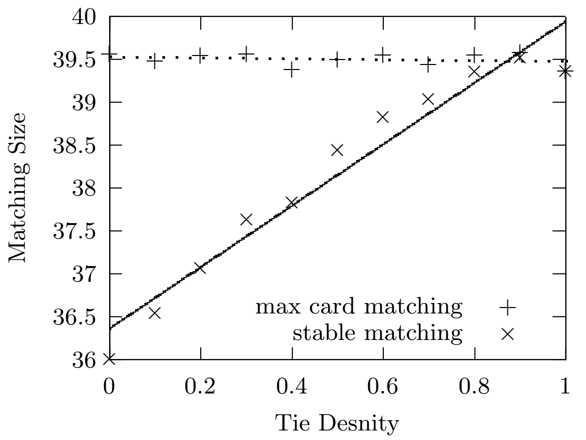

Figure 5: ***Average size of*** MAX-SMTI ***and*** MAX-CARD SMTI ***matchings, varying tie density***

ings and the number of blocking agents in MAX-CARD MIN-BA matchings. In all 4 cases we saw a very strong linear relationship between the size of the instance and the other aspects we are measuring. From this we may conject that results on the number of both blocking pairs and agents is proportional to the size of instances, and results on the nature of these are independent of the size of the instances.

The next set of experiments varied the tie density for men and women independently. We present the results from symmetrically varying the tie density in in Figures 5, 6 and 7. We saw that, besides a small amount of noise due to the number of instances, we have in general that the difference in cardinality of a stable and maximum cardinality matching decreases with the number of ties on the side with less ties. We can see that for both the number of blocking pairs and agents we have decrease that is inversely proportional to the tie density.

In Figures 8, 9 and 10 we present our results from symmetrically varying the lengths of preference lists for both the men and women, ignoring the case where women have

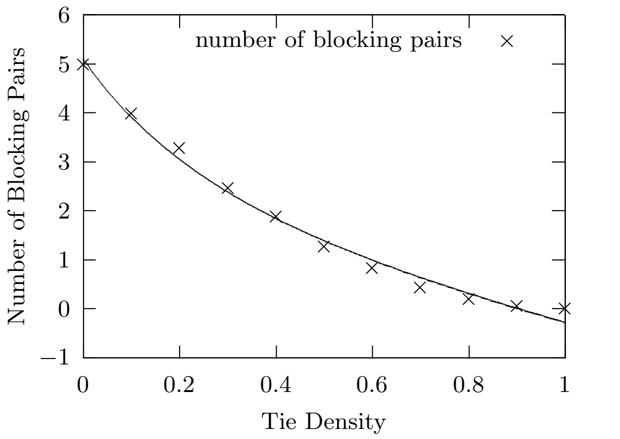

Figure 6: ***Average number of blocking pairs in a*** MAX-CARD MIN-BP SMTI ***matching, varying tie density***

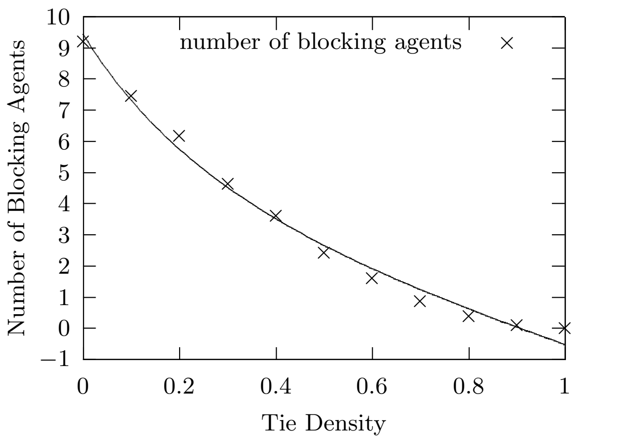

Figure 7: ***Average number of blocking agents in a*** MAX-CARD MIN-BA SMTI ***matching, varying tie density***

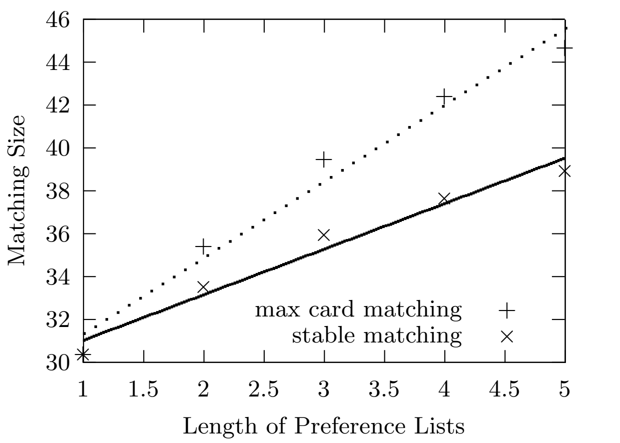

Figure 8: ***Average size of*** SMI ***and*** MAX-CARD SMI***, varying preference lists' length***

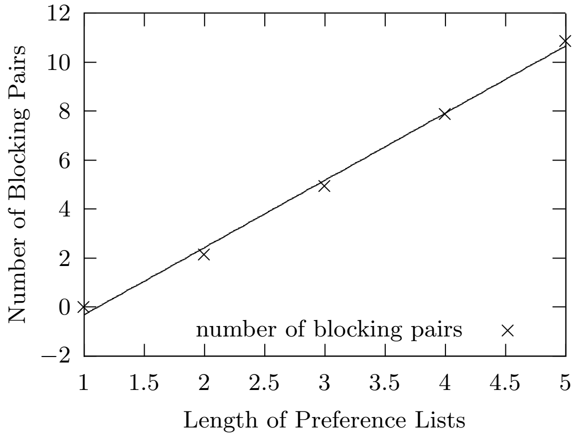

Figure 9: ***Average number of blocking pairs in a*** MAX-CARD MIN-BP SMI***, varying preference lists' length***

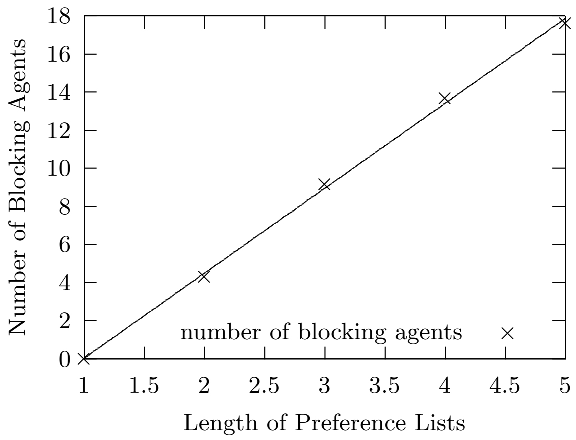

Figure 10: ***Average number of blocking agents in a*** MAX-CARD MIN-BA SMI***, varying preference lists' length***

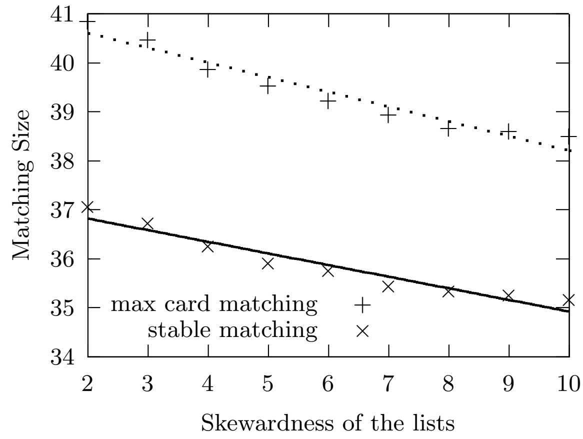

Figure 11: ***Average size of*** SMI ***and*** MAX-CARD SMI ***matchings, varying the skewedness of preference lists***

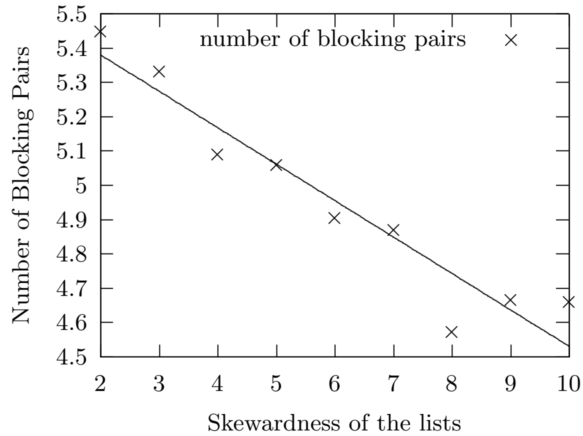

Figure 12: ***Average number of blocking pairs in*** MAX-CARD MIN-BP SMI***, varying the skewedness of preference lists***

smaller preference list due to the symmetry of the problem. In this we again see that, similar to the case with tie densities, the difference in cardinality between stable and maximum cardinality matchings, as well as the number of blocking pairs and agents in MAX-CARD MIN-BP and MAX-CARD MIN-BA matchings respectively, are all seemingly only impacted by the length of the shorter of the preference list.

In our final SMI experiment we looked at varying the skewedness for both the men and the women independently, showing the symmetrical results in Figures 11, 12 and 13. In these we found a weak but consistent correlation between the skewedness of the preference lists and the size of matchings, as well as the number of blocking pairs in a MAX-CARD MIN-BP and blocking agents in MAX-CARD MIN-BA matchings. In this case we see a slight decrease across all three values as the skewedness increases. Unlike in our experiments varying tie density and length of preference lists the impact seems based on the skewedness of both lists, as opposed to just from the more extreme one.

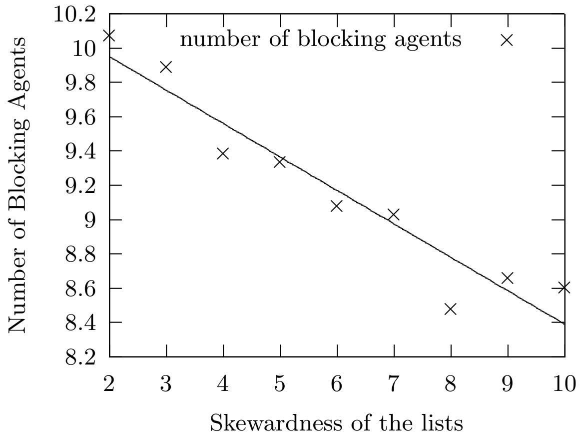

Figure 13: ***Average number of blocking agents in*** MAX-CARD MIN-BA SMI***, varying the skewedness of preference lists***

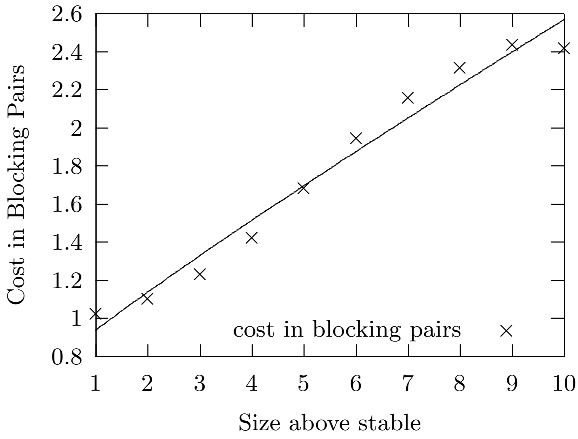

Figure 14: ***Cost in blocking pairs as we move further from a stable matching***

6.2 Optimal solutions for HRT

We will now consider our optimal solutions for the out HRT instances, presented in the same order as in the previous section. In this section all figures may be found in Appendix D.

We will again look at our experiments for varying the size of instances first. We can see the in these experiments that we had larger matchings on average, with more blocking pairs and agents, than the corresponding matchings for SMI of the same size. We can understand this by considering that the average resident's preference list should provide 18 posts that they find acceptable compared to the 3 available in the SMI case. A second reason is that as there is no upper bound on the lengths of hospitals lists, unlike for men, women and residents, the instances we generate will have longer preference lists on average.

For tie density we see again that in general increase the tie density for either list decreases the difference between the size of stable and maximum cardinality matchings, as well as the number of blocking pairs and agents in MAX-CARD MIN-BP and MAX-CARD MIN-BA matchings respectively. We see a much larger impact when varying the tie density on residents lists than for hospitals. One potential reason for this is that we have more residents than hospitals, meaning an increase in tie density for residents can affect many more lists than for hospitals. A second potential reason is that a tie on the resident's list will tie on average 12 posts, where as hospitals will only tie two residents.

For the lengths of preference lists we see a much larger impact on all three factors from the lengths of hospital's lists than from residents. For the difference in cardinality of matchings we see a strong increase in the difference between stable and maximum cardinality matchings as we increase the lengths of hospitals preference lists, with the difference seemingly being maximised when residents and hospitals have lists of equivalent length. However our current data is not sufficient to draw any significant conclusions on that. For both blocking pairs and agents we again see a strong increase in the number in a MAX-CARD MIN-BP and MAX-CARD MIN-BA matching as we increase the length of the hospitals lists, while seeing in general a small decrease as we increase the lengths of residents preference lists. The relative importance of the hospitals lists makes intuitive sense, as they are able to match multiple residents an increase of preference lists by 1 until we reach the capacity of the lists will have a much larger impact by increasing the matching. Unfortunately we don't have sufficient data to draw any significant conclusions on the impact of residents lists, however it is certainly something worth investigating further based on these results.

Finally we see in the skewedness case much more impact again from the hospitals preference lists being skewed than the resident's. In the cardinality case in particular we see relatively little impact on the skewedness of residents lists on the cardinality of matchings. In the blocking pair and to a lesser extent blocking agent cases we do see some impact, with a general increase in the number of blocking pairs or agents corresponding to an increase in the skewedness of the residents list. It is particularly interesting that this is more pronounced in the blocking pair case than the blocking agent one as in our other experiments we have in general seen very close results between the two. This suggests that we are finding instances where either the MAX-CARD MIN-BP and MAX-CARD MIN-BA matchings are different, or that in those matchings we have cases of most blocking agents forming blocking pairs with a small set of popular blocking agents.

One notable result we found in both our SMTI and HRT experiments was in regards to the cost in both blocking pairs and blocking agents of finding a matching as we increase the distance from a stable matching. More formally, we found that given an instance admitting a stable matching of cardinality N, the cost from moving from a matching of size $N+T$ to one of $N+T+1$ was in general less than the cost of moving from a matching of size $N+T+1$ to one of size $N+T+2$, for both blocking pairs and agents relative to whichever we were trying to minimise. We illustrate this in Figures 14 and 15 for our experiments for SMTI MAX-CARD MIN-BP and SMTI MAX-CARD MIN-BA respectively for our instances of size 100, up to an increase of 10 beyond which there were to few instance to draw any significant conclusions. In our experiments we saw that the increase for both

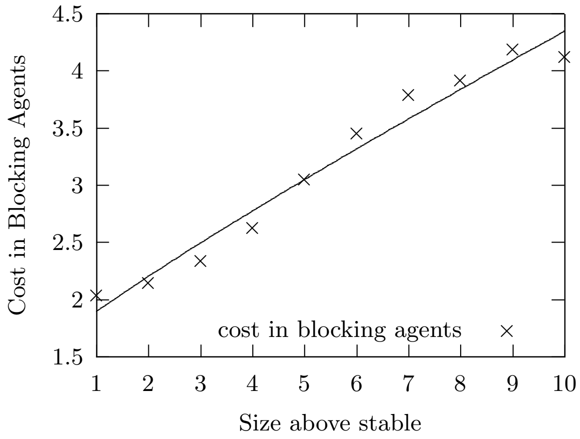

Figure 15: *Cost in blocking agents as we move further from a stable matching*

m_1:	$\{w_1\ w_2\}$	w_1:	$\{m_1\ m_2\}$
m_2:	$\{w_1\}$	w_2:	$\{m_2\}$
m_3:	$\{w_3\ w_4\}$	w_3:	$\{m_3\ m_4\}$
m_4:	$\{w_3\}$	w_4:	$\{m_3\}$
m_5:	$\{w_5\ w_6\}$	w_5:	$\{m_5\ m_6\}$
m_6:	$\{w_5\}$	w_6:	$\{m_5\}$

Table 1: *Example instance with constant cost to increase cardinality*

was in general linear with blocking pairs increasing slightly faster relative to the lower bound (1) than blocking agents did relative to the their lower bound (2). We would assume from this that as we tend towards a maximum cardinality matching, the average rank of the agents that are matched increases, causing more blocking pairs and agents to emerge within the matchings. We can see that this is not a strict rule, considering the instance in Table 1.

In this case we can find a stable matching of size 3, $\{(m_1, w_1), (m_3, w_3), (m_5, w_5)\}$, as well as a matching of cardinality 4 containing 1 blocking pair, $\{(m_1, w_2), (m_2, w_1), (m_3, w_3), (m_5, w_5)\}$, one of cardinality 5 containing 2 blocking pairs, $\{(m_1, w_2), (m_2, w_1), (m_3, w_4), (m_4, w_3), (m_5, w_5)\}$ and finally one of size 6 with 3 blocking pairs, $\{(m_1, w_2), (m_2, w_1), (m_3, w_4), (m_4, w_3), (m_5, w_6), (m_6, w_5)\}$. In this case we have that the presented matchings are both minimal in terms of blocking pairs and blocking agents, and have a constant cost in both cases as the lower bound.

6.3 Real Instances

To analyse the effectiveness of our model we looked at the instances for the Scottish Foundation Allocation Scheme in 2006, 2007 and 2008. These instances are much larger than the instances we experimented on, having over 700 residents and 50 hospitals in all 3. We also had longer preference lists, with a bound of 6 on the residents lists and unbounded lists for the hospitals. We also had a large number of ties on the hospitals lists, with ties containing in many cases over 10 residents.

A matching of maximum cardinality for the 2007 instance was found by Manlove et al. [4], admitting 400 blocking

Instance	|R|	|H|	|P|	|S|	|M|	|BP(M)|
2006	759	53	801	758	759	8
2007	781	53	789	746	781	150
2008	748	52	752	709	745	143

Table 2: *Real* HRT *instances*

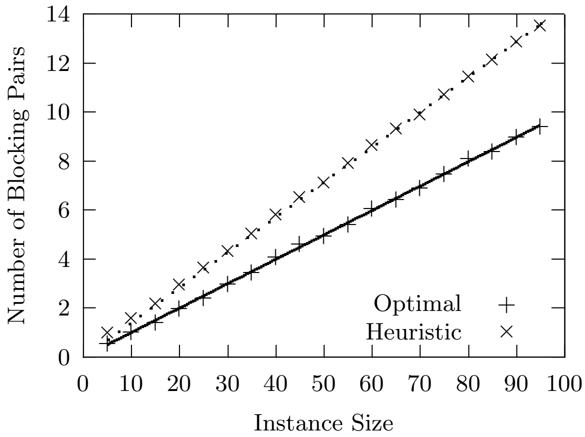

Figure 16: *Number of blocking pairs found by the Heuristic algorithm and the optimal case, varied over instance size for* SMI

pairs. Our results for the same instance provided a significant increase admitting only 150 blocking pairs. It is worth noting that the matchings we found for all three instances were not necessarily optimal, but rather just the best ones we could find. We show these results in the following table, using |R| to denote the number of residents, |H| the number of hospitals, |P| the number of posts, |S| the size of a stable matching, |M| the size of a maximum cardinality matching, and |BP(M)| smallest number of blocking pairs. We show our results in Table 2.

The first thing we can see is that these matchings, though not necessarily optimal, seem to have a similar amount of blocking pairs as we might expect from our experimental results, although for the 2006 case it seems likely that we may find a matching with fewer blocking pairs. Beyond just the results, we may also note that our instances in this case are much larger than for our experiments, meaning that while we were not able to find optimal matchings, that we were able to find what seems to be relatively good matchings suggest that our model may be able to be used for real world applications.

6.4 Heuristic Results

Here we will look at the use of our heuristic for finding matchings with a minimum number of blocking pairs. To test this we used our data sets for varying cardinality, for both SMTI and HRT instances, comparing the best result we got from a heuristic to the optimal solutions we previously computed. We present these results in Figures 16 and 17.

As with the optimal case, we see a linear relationship between the number of blocking pairs and the size of the instance. As we would expect from this, we have the number of blocking pairs increasing by a larger factor as we increase

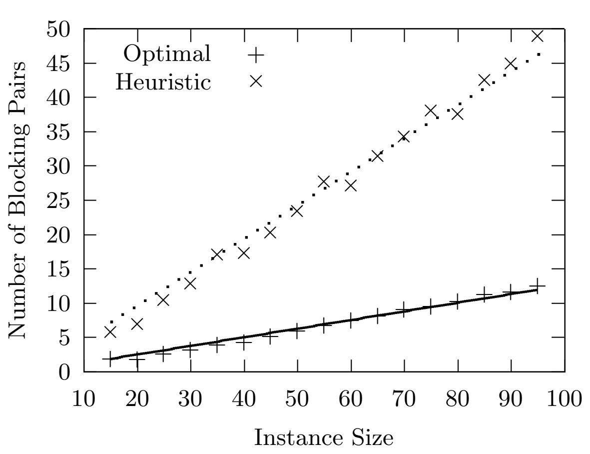

Figure 17: ***Number of blocking pairs found by the Heuristic algorithm and the optimal case, varied over instance size for*** HR

Weighting	SMI		HR		total	
	alone	total	alone	total	alone	total
1	5955	19297	9229	11286	15184	30583
2	703	14045	6714	8769	7417	22814
3	0	1058	0	442	0	1500
4	0	1058	0	442	0	1500

Table 3: ***Number of times each weighting found the smallest number of blocking pairs, either as the only one returning that result or alongside others***

cardinality than the optimal case. We would assume this, as it becomes more likely that we encounter a locally optimal solution with more blocking pairs increases. Due to time constraints we were unable to preform experiments beyond where we varied the size of instances, however we would expect to see the same relationship in between the blocking pairs in the optimal solution and the solutions we computed.

Looking at the relative performances of the heuristic weightings described in Section 4, we found that the first weighting found the optimal solution far more often than the others, however in the majority of cases this was tied with the second weighting. In all cases the third and fourth weighting computed the worst solution. We do see a notable disparity between SMI and HR, with weightings 1 and 2 being much closer for HR than SMI, while weightings 3 and 4 are worse in HR than SMI. We also see that the algorithms preformed notably worse for HR, having a significantly higher factor between the optimal number of blocking pairs and the number found by our algorithm than the SMI case.

It is interesting to see the large disparity between the first weighting and the second, which we would assume to be similar in many cases due to the lengths of preference lists being very restricted. Looking at the number of times these heuristics returned an optimal solution we see an exponential decrease, which we show in Figure 18. Here we see an exponential decrease in the number of optimal cases we find solutions to as we increase cardinality, with the increase being far more rapid for weightings 3 and 4 as we would expect based on the previous results. In the case for HR we have

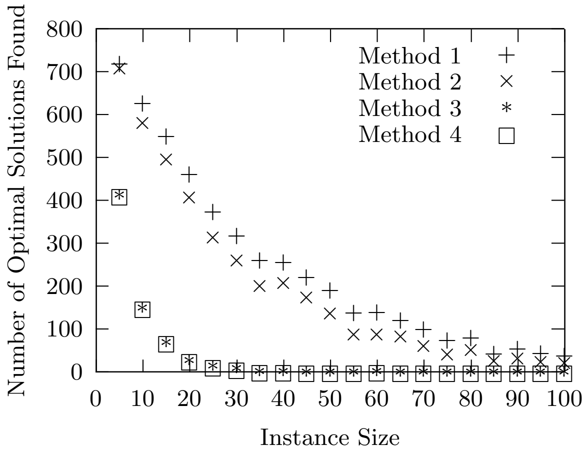

Figure 18: ***Number of optimal solutions versus size of instances for*** SMI

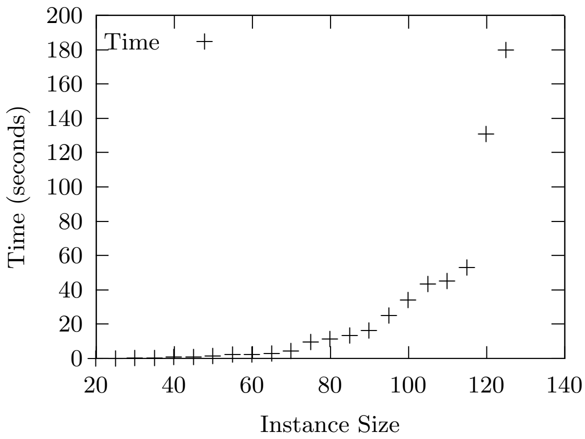

Figure 19: ***Average time to solve instances of*** MAX-CARD MIN-BP HR ***as we vary the size of instances***

much fewer optimal matchings, with a mere 34 out of 1000 in the best case, and a similar rate of decrease. This combined with the poor results from the algorithm in general suggests that we may need to investigate other techniques for finding heuristics to HR.

6.5 Scalability and use of other solvers

Finally we will look at the scalability of instance size and the time taken to solve to optimality. All experiments were done on a machine with 8GB of RAM, and a AMD 7th Gen A10-9600P APU with 4 cpu cores and 2.6 GHz clock speed. The instances used were HR instances with the same settings as discussed in Section 5.

We see that our IP model seems to scale exponentially with the size of instances. We may assume that the sudden jump between 115 and 120 in our samples were to do with the size of instances requiring paging of some kind to store the models. These results largely conform to our expectations, with the slight irregularities in the graph curve presumably due to the limited number of instances we were able to experiment on.

A MiniZinc encoding was made, however even for relatively small instances of a size of 20 this encoding struggled to find solutions some instances within a reasonable time frame. It is worth noting that this encoding took a far greater amount of time to find solutions to instances where the maximum cardinality was the same as stable cardinality instance. This suggests that we may be able to take advantage of some degree of pre-processing, providing a lower bound on the number of blocking pairs for more complex cases, that may result in a speed up, however we were not able to get such an encoding working adequately within the time frame of the project.

7. CONCLUSION AND FURTHER WORK

In this paper we have provided new theoretical bounds on the numbers of blocking pairs and agents in MAX-CARD MIN-BP SMTI and MAX-CARD MIN-BA SMTI matchings. In particular we showed that as we increase the cardinality of a matching above stable, we must also increase the number of blocking pairs or agents respectively. We also provided new structural results for this, showing that any blocking pairs must only be between agents that are matched in any above stable matching where we are minimising the number of blocking pairs or blocking agents.

Our investigations using IP models showed that in general there seems to exists a linear relationship between the size of an instance and the cardinality of both stable and maximum matchings, as well as the number of blocking pairs and blocking agents in a MAX-CARD MIN-BP SMTI or MAX-CARD MIN-BA SMTI matching respectively (as well as for the case for these problems with HRT instances). We also saw that the relationships for varying tie density, lengths of preference lists and skewedness of preference lists have clear, though not necessarily linear, relationships for all three factors.

In the case of SMTI we saw that for tie density the impact was based on the side with lower tie density. Similarly for preference lists the impact was strongly correlated with the length of the smaller preference lists. In skewedness there was a weak but noticeable impact based on the skewedness of both lists. The results were very similar for HRT, though we did see that more impact was made by the residents in the case of varying tie densities and by the hospitals in cases of the lengths and skewedness of preference lists. We also looked at some real instances, finding a considerable improvement over the previous best case where available. This also showed that our tools were able to scale to handling lager and more complicated instances.

Beyond just the IP models we also looked at a heuristic for MAX-CARD MIN-BP SMTI and MAX-CARD MIN-BP HRT, based on our structural results. This heuristic proved to be considerably more effective for SMTI than HRT, although still scaling linearly with the sizes of instances in both cases. Finally we looked at the scalability of the IP solver, showing an exponential increase in the time to solve instances as size increased.

7.1 Further Work

The first major piece of further work would be to extend the results presented in Section 2 to HRT. While it seems likely that these should hold, further analysis is required to ensure they do in the many-to-one generalisation. It may also be worth investigating if these results have any baring on our understanding of similar problems such as stable roommates.

One major open question raised by our empirical results is to see if there is any correlation between MAX-CARD MIN-BA and MAX-CARD MIN-BP matchings. While there clearly appears to be looking at our results, this may be due to the size of instances we are considering and certainly requires further investigation. Another related problem would be the structure of matchings that are below stable cardinality, in particular seeing if, in such matchings, the number of blocking pairs decreases as the cardinality of these matching increases towards stable. Also an investigation of the structures of these matchings, particularly if one or more of the types of blocking pairs can not exist or must exist in these matchings.

Adapting our model for other problems, such as the student project allocation problem or the stable roommates problem, is also a direction that we may wish to take. Given the improvements we've seen in the results to the real instances of HRT this may show similar improvements to these problems. Attempts to develop heuristics to focus on MAX-CARD MIN-BA SMTI, as well as to investigate if our heuristics for MAX-CARD MIN-BP SMTI were effective for that problem. Further work on using techniques such as finding augmenting paths from a stable matching, taking advantage of the theoretical results such as being able to resolve bridge blocking pairs and internal cycles to develop a more effective heuristic.

Finally, attempting to adapt our models to be better aimed at constraint programming and Boolean satisfaction problem solvers. This may allow us to find solutions faster than with the IP solver, or at least to provide a good comparison between the solvers. One direction we may take this in would be providing bounds on the number of blocking pairs to avoid having to search for matchings with an impossible number of blocking pairs. We may also combine this with an upper-bound from a heuristic to provide a starting matching.

Acknowledgements

I would like to thank Dr. David Manlove for his continued help and support throughout the project. I would also like to thank Dr. William Patterson for his comments on the proof in Section 2

8. REFERENCES

[1] D.J. Abraham, P. Biro, and D.F. Manlove. "Almost Stable" Matchings in the Roommates problem. *Proceedings of WAOA 2005: the 3rd International Workshop on Approximation and Online Algorithms*, 3879:1 – 14, 2006.

[2] David L. Applegate, Robert E. Bixby, VaÅąek ChvÃątal, William Cook, Daniel G. Espinoza, Marcos Goycoolea, and Keld Helsgaun. Certification of an optimal TSP tour through 85,900 cities. *Operations Research Letters"*, 37(1):11 – 15, 2009.

[3] Péter Biró, David F. Manlove, and Iain McBride. The Hospitals / Residents Problem with Couples: Complexity and Integer Programming Models. In Joachim Gudmundsson and Jyrki Katajainen, editors, *Experimental Algorithms: 13th International Symposium, SEA 2014, Copenhagen, Denmark, June*

29 – July 1, 2014. Proceedings, pages 10–21. Springer International Publishing, Cham, 2014.

[4] Péter Biró, David F. Manlove, and Shubham Mittal. Size versus Stability in the marriage problem. *Theoretical Computer Science*, 411:1828 – 1842, Febuary 2009.

[5] G Dantzig, R Fulkerson, and S Johnson. Solution of a large-scale traveling-salesman problem. *Operations Research*, 2:393–410, 1954.

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

[7] D. Gale and M. Sotomayor. Some remarks on the stable mariage problem. *Discrete Applied Mathematics*, 11:223–232, 1985.

[8] Martin Grötschel. On the symmetric travelling salesman problem: solution of a 120-city problem. *Mathematical Programming Study*, 12:61–77, 1980.

[9] Dan Gusfield and Robert W. Irving. *The Stable Marriage Problem Structure and Algorithms*. The MIT Press, 1989.

[10] Robert W. Irving, David F. Manlove, and Sandy Scott. The Hospitals/Residents Problem with Ties. In *Algorithm Theory - SWAT 2000*, pages 259–271. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[11] Kazuo Iwama, Shuichi Miyazaki, Yasufumi Morita, and David Manlove. Stable Marriage with Incomplete Lists and Ties. In Jiří Wiedermann, Peter van Emde Boas, and Mogens Nielsen, editors, *Automata, Languages and Programming: 26th International Colloquium, ICALP'99 Prague, Czech Republic, July 11–15, 1999 Proceedings*, pages 443–452. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[12] Thorsten Koch, Tobias Achterberg, Erling Andersen, Oliver Bastert, Timo Berthold, Robert E. Bixby, Emilie Danna, Gerald Gamrath, Ambros M. Gleixner, Stefan Heinz, Andrea Lodi, Hans Mittelmann, Ted Ralphs, Domenico Salvagnin, Daniel E. Steffy, and Kati Wolter. MIPLIB 2010. *Mathematical Programming Computation*, 3(2):103, Jun 2011.

[13] Augustine Kwanashie and David F. Manlove. An Integer Programming Approach to the Hospitals/Residents Problem with Ties. In Dennis Huisman, Ilse Louwerse, and Albert P.M. Wagelmans, editors, *Operations Research Proceedings 2013: Selected Papers of the International Conference on Operations Research, OR2013, organized by the German Operations Research Society (GOR), the Dutch Society of Operations Research (NGB) and Erasmus University Rotterdam, September 3-6, 2013*, pages 263–269. Springer International Publishing, Cham, 2014.

[14] David F. Manlove, Robert W. Irving, Kazuo Iwama, Shuichi Miyazaki, and Yasufumi Morita. Hard Variants of Stable Marriage. *Theor. Comput. Sci.*, 276(1-2):261–279, April 2002.

[15] David F. Manlove and Iain Mcbride. The Hospitals / Residents problem with Couples: Complexity and Integer Programming models. *CoRR*, abs/1308.4534, 2013.

[16] David F. Manlove, Iain Mcbride, and James Trimble. "Almost-stable" Matchings in the Hospitals / Residents Problem with Couples. *Constraints*, 22(1):50–72, January 2017.

[17] D.F.; Manlove. Stable Marriage with Ties and Unacceptable Partners. Technical report, University of Glasgow, Department of Computing Science, 1999.

[18] Bernhard Meindl and Matthias Templ. Analysis of Commercial and Free and Open Source Solvers for the Cell Suppression Problem. *Trans. Data Privacy*, 6(2):147–159, August 2013.

[19] Shubham Mittal. Experiments on size versus stability. Technical report, University of Glasgow, 2006.

[20] A.E. Roth. On the allocation of residents to rural hospitals a general property of two-sided matching markets. *Journal of Economic Theory*, 54:425–427, 1986.

Appendices

A. EXTENSION TO MAX-CARD MIN-BP HRT

We will here see how this model may be extend to handle an HRT instance, instead of an SMTI one. In this case we have the preference matrices as before, as well as a capacity vector c, where c_i is the capacity of the i^{th} hospital. We will consider the number the residents to be n_1 and the number of hospitals to be n_2. As before we will have a dummy hospital to represent unmatched residents, and a dummy resident to represent undersubscribed hospitals.

Both objective functions will be unchanged with the extension to HRT. When we are evaluating the first objective function we will be subject to Constraints 2, 3 and 5, as well as the following two constraints:

$$\sum_{i=1}^{n_1+1} m_{i,j} \leq c_j \qquad (1 \leq j \leq n_2) \tag{8}$$

$$m_{n_1+1,j}c_j + \sum_{i=1}^{n_1} m_{i,j} \geq c_j \qquad (1 \leq j \leq n_2) \tag{9}$$

These require some intuition to understand that they will provide the appropriate values. Constraint 8 provides an upper bound on the amount of residents that may be assigned to the hospital as being equal to the capacity of the hospital. Constraint 9 ensures that if the capacity is not reached, the hospital is matched to the dummy resident.

When evaluating the second objective function, we will be subject to Constraints 2, 3 and 5 to 7 from the previous formulation, as well as Constraints 8 and 9. As we have already considered the use of Constraints 2, 3 and 5 on this model, we will show that that Constraint 7 will still be sufficient to provide the relevant blocking pair information. For any resident-hospital pair (r_i,h_j), resident r_i only has to be more preferable than h_j's least preferable resident to form a blocking pair. Clearly from the Constraint we will be testing against the least preferable resident that the hospital is assigned to, including the dummy resident in the case that it is undersubscribed. From the arguments in Section 3, can see that the when (r_i,h_j) form a blocking pair, we will have a positive strictly between 1 and 0 on the right hand side of the inequality, forcing the value of $b_{i,j}$ to be 1. In the case that they do not, the value of the right hand side will be less than 0. We can also see that even though the right hand side may be less than 0 when evaluating against assigned residents rather than the least preferable will not effect the final value of $b_{i,j}$.

It is worth noting that, as SMTI is a special case of HRT we can use this model to solve a SMTI instance by constraining the values of the capacity vector to be 1 for all hospitals. We can see that substituting c_j with 1 in constraints 8 and 9, we get two constraints that are equivalent to that of constraint 4, as we would expect.

B. MAXIMUM STABLE MATCHINGS

In this section we will look at changes we may make to the model in Section 3 to find matchings for both HRT and SMTI that are the maximum possible stable matchings.

For all these cases we will assume that we are covering the case of SMTI, however as noted in Appendix A, we can trivially extend this to HRT. We will also see in Appendix C how we may extend these to cover the case of blocking agents.

Our objective function for*maximum cardinality stable matching* (MAX-SMTI/HRT) is to maximise the value of $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} m_{i,j}$ with Constraints 2 to 5, alongside the additional constraint:

$$1 \leq i \leq n_1, 1 \leq j \leq n_2, 1 \leq q \leq n_2+1, 1 \leq p \leq n_1+1$$
$$\frac{(P^U_{i,j} - P^U_{i,q})m_{i,q}}{|(P^U_{i,j} - P^U_{i,q})|+1} + \frac{(P^W_{i,j} - P^W_{p,j})m_{p,j}}{|(P^W_{i,j} - P^W_{p,j})|+1} \leq \frac{n_1+n_2}{n_1+n_2+1} \tag{10}$$

This ensures that we can have no blocking pairs, which we can see trivially from the arguments from Section 3.

C. EXTENSION OF THE MODELS TO COVER BLOCKING AGENTS

In this section we will see how we might convert our model for finding the minimum number of blocking pairs to instead find the minimum number of blocking agents. We may observe that it is sufficient to simply add two vectors, V^U and V^W, constrained to binary variables, such that:

$$V^U_i = \begin{cases} 1, & \text{if } m_i \text{ is a blocking agent} \\ 0, & \text{otherwise} \end{cases}$$

$$V^W_j = \begin{cases} 1, & \text{if } w_j \text{ is a blocking agent} \\ 0, & \text{otherwise} \end{cases}$$

By adding these alongside the following constraints:

$$V^U_i \geq b_{i,j} \qquad 1 \leq i \leq n_1, 1 \leq j \leq n_2 \tag{11}$$
$$V^W_j \geq b_{i,j} \qquad 1 \leq i \leq n_1, 1 \leq j \leq n_2 \tag{12}$$

With the objective to minimise the value of $\sum_{i=1}^{n_1} V^U_i + \sum_{j=1}^{n_2} V^W_j$.

We can reduce the number of variables that we are using in this by replacing Constraint 7 with the following:

$$1 \leq i \leq n_1, 1 \leq j \leq n_2, 1 \leq q \leq n_2+1, 1 \leq p \leq n_1+1$$
$$V^U_i \geq \frac{(P^U_{i,j} - P^U_{i,q})m_{i,q}}{|(P^U_{i,j} - P^U_{i,q})|+1} + \frac{(P^W_{i,j} - P^W_{p,j})m_{p,j}}{|(P^W_{i,j} - P^W_{p,j})|+1} - \frac{n_1+n_2}{n_1+n_2+1} \tag{13}$$

$$1 \leq i \leq n_1, 1 \leq j \leq n_2, 1 \leq q \leq n_2+1, 1 \leq p \leq n_1+1$$
$$V^W_j \geq \frac{(P^U_{i,j} - P^U_{i,q})m_{i,q}}{|(P^U_{i,j} - P^U_{i,q})|+1} + \frac{(P^W_{i,j} - P^W_{p,j})m_{p,j}}{|(P^W_{i,j} - P^W_{p,j})|+1} - \frac{n_1+n_2}{n_1+n_2+1} \tag{14}$$

Alleviating the need for Constraints 11 and 12. We can see from the arguments in Section 3 and A that these will be 1 if and only if the agent is in a blocking pair. Again, minimising the same function we will get the minimum number of blocking pairs in the instance. We can also see how we could combine these Constraints with the others we discussed for the problems in Appendix B to investigate the properties for these matchings.

In order to find a matching with a minimum number of blocking agents from the matching with a set cardinality with the minimum number of blocking pairs (K-CARD MIN-BP MIN-BA), we would evaluate the model from Appendix B with K set to the cardinality we are searching for, getting B blocking pairs. We would then attempt to minimise the value of $\sum_{i=1}^{n_1} V_i^U + \sum_{j=1}^{n_2} V_j^W$, subject to the relevant previous Constraints from Appendix B, alongside the new Constraint:

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} b_{i,j} = B \tag{15}$$

Which will ensure that we get the minimum number of blocking pairs. We can trivially see how we might change this model to find a matching with a minimum number of blocking pairs from the matching with a set cardinality with the minimum number of blocking agents (K-CARD MIN-BA MIN-BP).

D. HRT FIGURES

Please note that in Figures 23 - 31 we are using a heat map where a lighter shading indicates a higher value.

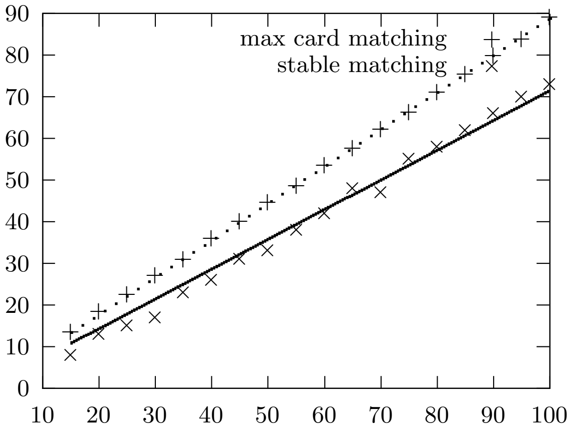

Figure 20: ***Instance Size versus Average Matching Size for stable and maximum cardinality matchings***

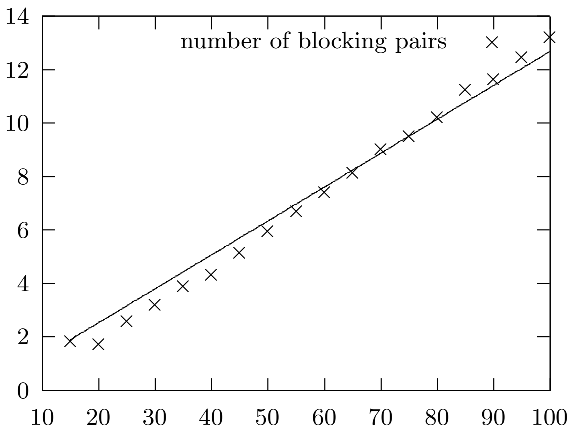

Figure 21: ***Instance Size versus Number of Blocking Pairs in a*** MAX-CARD MIN-BP ***matching***

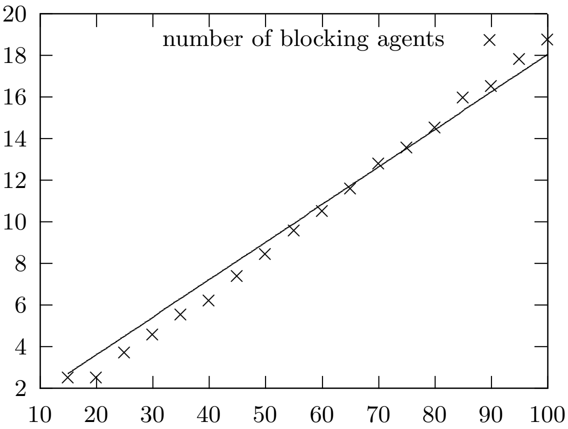

Figure 22: ***Instance Size versus Number of Blocking Agents in a*** MAX-CARD MIN-BA ***matching***

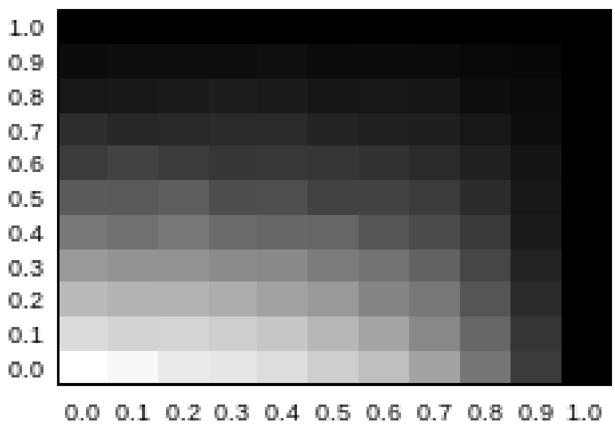

Figure 23: ***Difference in cardinality of a*** MAX-HRT ***and*** MAX-CARD HRT ***as we vary the tie density***

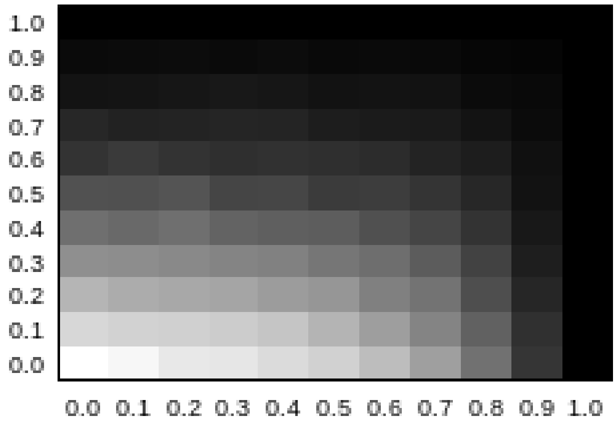

Figure 24: *Number of blocking pairs in* MAX-CARD MIN-BP *as we vary the tie density*

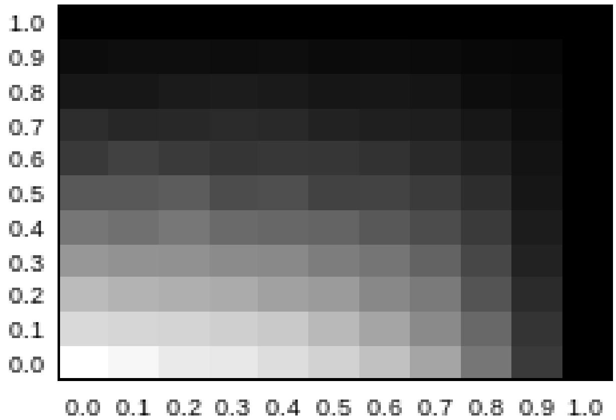

Figure 25: *Number of blocking agents in* MAX-CARD MIN-BA *as we vary the tie density*

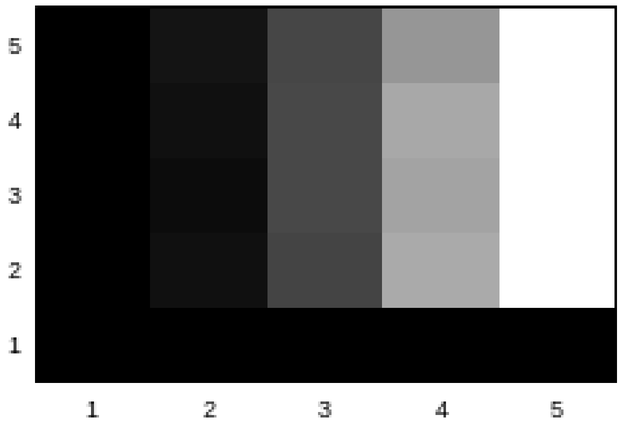

Figure 26: *Difference in cardinality of a* MAX-HRT *and* MAX-CARD HR *as we vary the length of preference lists*

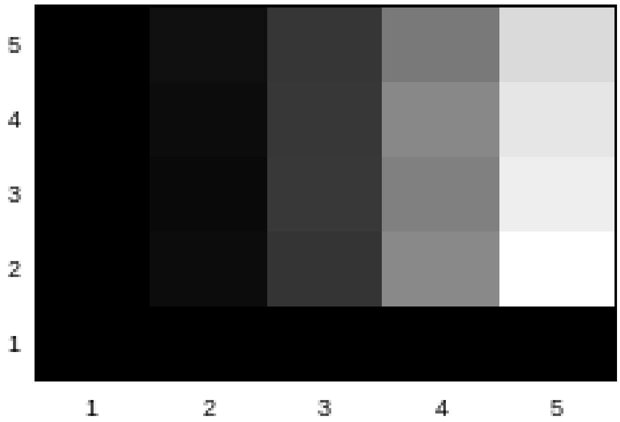

Figure 27: *Number of blocking pairs in* MAX-CARD MIN-BP *as we vary the length of preference lists*

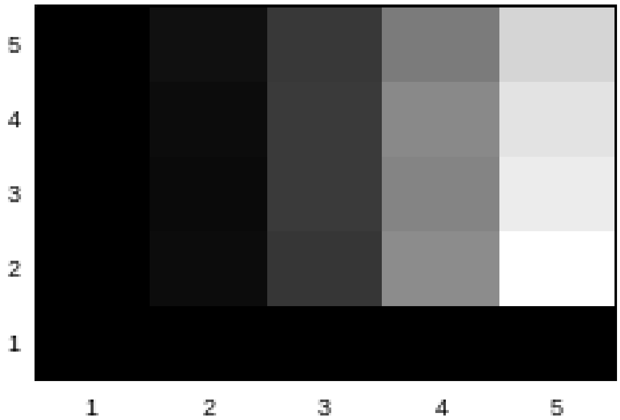

Figure 28: *Number of blocking agents in* MAX-CARD MIN-BA *as we vary the length of preference lists*

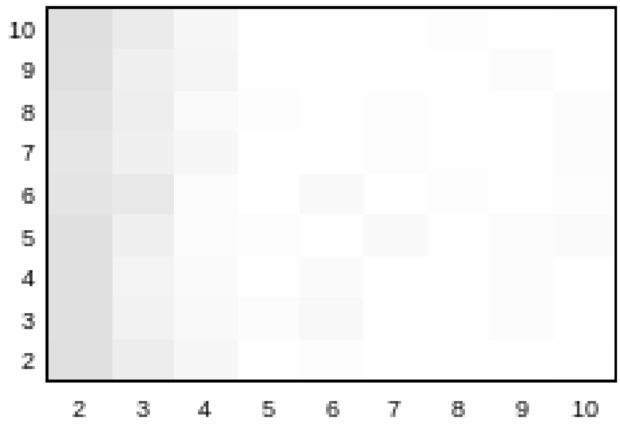

Figure 29: *Difference in cardinality of* HR *and* MAX-CARD HR *as we vary the skewedness*

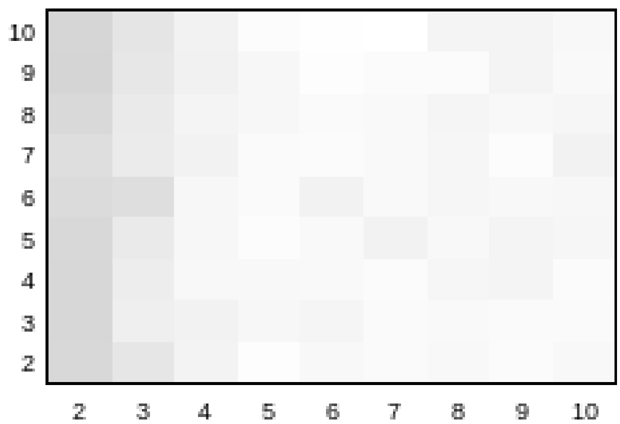

Figure 30: *Number of blocking pairs in* MAX-CARD MIN-BP *as we vary the skewedness*

Figure 31: *Number of blocking agents in* MAX-CARD MIN-BA *as we vary the skewedness*