A $\frac{3}{2}$-approximation algorithm for the Student-Project Allocation problem

Frances Cooper
School of Computing Science, University of Glasgow
[Glasgow, Scotland, UK]
frances.cooper.1@research.gla.ac.uk
[https://orcid.org/0000-0001-6363-9002]

David Manlove
School of Computing Science, University of Glasgow
[Glasgow, Scotland, UK]
david.manlove@glasgow.ac.uk
[https://orcid.org/0000-0001-6754-7308]

Abstract

The Student-Project Allocation problem with lecturer preferences over Students (SPA-S) comprises three sets of agents, namely students, projects and lecturers, where students have preferences over projects and lecturers have preferences over students. In this scenario we seek a stable matching, that is, an assignment of students to projects such that there is no student and lecturer who have an incentive to deviate from their assignee/s. We study SPA-ST, the extension of SPA-S in which the preference lists of students and lecturers need not be strictly ordered, and may contain ties. In this scenario, stable matchings may be of different sizes, and it is known that MAX SPA-ST, the problem of finding a maximum stable matching in SPA-ST is NP-hard. We present a linear-time $\frac{3}{2}$-approximation algorithm for MAX SPA-ST and an Integer Programming (IP) model to solve MAX SPA-ST optimally. We compare the approximation algorithm with the IP model experimentally using randomly-generated data. We find that the performance of the approximation algorithm easily surpassed the $\frac{3}{2}$ bound, constructing a stable matching within 92% of optimal in all cases, with the percentage being far higher for many instances.

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms

Keywords and phrases Matching problems, Approximation, Algorithms, Stability

Digital Object Identifier 10.4230/LIPIcs.SEA.2018.8

1 Introduction

Background and motivation. In universities all over the world, students need to be assigned to projects as part of their degree programmes. Lecturers typically offer a range of projects, and students may rank a subset of the available projects in preference order. Lecturers may have preferences over students, or over the projects they offer, or they may not have explicit preferences at all. There may also be capacity constraints on the maximum numbers of students that can be allocated to each project and lecturer. The problem of allocating students to projects subject to these preference and capacity constraints is called the Student-Project Allocation problem (SPA) [Section 5.5] [4, 5]. Variants of this problem can be defined for the cases that lecturers have preferences over the students that rank their
A \( \frac{3}{2} \)-approximation algorithm for the Student-Project Allocation problem

projects [2], or over the projects they offer [10], or not at all [7]. In this paper we focus on the first of these cases, where lecturers have preferences over students – the so-called Student-Project Allocation problem with lecturer preferences over Students (SPA-S).

Finding an optimal allocation of students to projects manually is time-consuming and error-prone. Consequently many universities automate the allocation process using a centralised algorithm. Given the typical sizes of problem instances (e.g., 130 students at the University of Glasgow, School of Computing Science), the efficiency of the matching algorithm is of paramount importance. In the case of SPA-S the desired matching must be stable with respect to the given preference lists, meaning that no student and lecturer have an incentive to deviate from the given allocation and form an assignment with one another [11].

Abraham et al. [2] described a linear-time algorithm to find a stable matching in an instance \( I \) of SPA-S when all preference lists in \( I \) are strictly ordered. They also showed that, under this condition, all stable matchings in \( I \) are of the same size. In this paper we focus on the variant of SPA-S in which preference lists of students and lecturers can contain ties, which we refer to as the Student-Project Allocation problem with lecturer preferences over Students including Ties (SPA-ST). Ties allow both students and lecturers to express indifference in their preference lists (in practice, for example, lecturers may be unable to distinguish between certain groups of students). A stable matching in an instance of SPA-ST can be found in linear time by breaking the ties arbitrarily and using the algorithm of Abraham et al. [2].

The Stable Marriage problem with Ties and Incomplete lists (SMTI) is a special case of SPA-ST in which each project and lecturer has capacity 1, and each lecturer offers one project. Given an instance of SMTI, it is known that stable matchings can have different sizes [9], and thus the same is true for SPA-ST. Yet in practical applications it is desirable to match as many students to projects as possible. This motivates the problem of finding a maximum (cardinality) stable matching in an instance of SPA-ST. This problem is NP-hard, since the corresponding optimisation problem restricted to SMTI which we refer to as MAX SMTI is NP-hard [9]. Király [6] described a \( \frac{3}{2} \)-approximation algorithm for MAX SMTI. He also showed how to extend this algorithm to the case of the Hospitals-Residents problem with Ties (HRT), where HRT is the special case of SPA-ST in which each lecturer \( l \) offers one project \( p \), and the capacities of \( l \) and \( p \) are equal. Yanagisawa [12] showed that MAX SMTI is not approximable within a factor of \( \frac{25}{29} \) unless P=NP; the same bound applies to MAX SPA-ST.

Our contribution. In this paper we describe a linear-time \( \frac{3}{2} \)-approximation algorithm for MAX SPA-ST. This algorithm is a non-trivial extension of Király’s approximation algorithm for HRT as mentioned above. We also describe an Integer Programming (IP) model to solve MAX SPA-ST optimally. Through a series of experiments on randomly-generated data, we then compare the sizes of stable matchings output by our approximation algorithm with the sizes of optimal solutions obtained from our IP model. Our main finding is that the performance of the approximation algorithm easily surpassed the \( \frac{3}{2} \) bound on the generated instances, constructing a stable matching within 92% of optimal in all cases, with the percentage being far higher for many instances.

Note that a natural “cloning” technique, involving transforming an instance \( I \) of SPA-ST into an instance \( I' \) of SMTI and then using Király’s \( \frac{3}{2} \)-approximation algorithm for SMTI [6] in order to obtain a similar approximation in SPA-ST does not work in general, as we show in Appendix A. This motivates the need for a bespoke algorithm for the SPA-ST case.

Structure of this paper. Section 2 gives a formal definition of SPA-ST. Section 3 describes the \( \frac{3}{2} \)-approximation algorithm, and the IP model for MAX SPA-ST is given in Section 4. The experimental evaluation is described in Section 5, and Section 6 discusses future work.
2 Formal definition of SPA-ST

An instance $I$ of SPA-ST comprises a set $S = \{s_1, s_2, ..., s_n\}$ of students, a set $P = \{p_1, p_2, ..., p_m\}$ of projects, and a set $L = \{l_1, l_2, ..., l_k\}$ of lecturers. Each project is offered by one lecturer, and each lecturer $l_k$ offers a set of projects $P_k \subseteq P$, where $P_1, ..., P_k$ partitions $P$. Each project $p_j \in P$ has a capacity $c_j \in \mathbb{Z}_{\geq 0}^+$, and similarly each lecturer $l_k \in L$ has a capacity $d_k \in \mathbb{Z}_{\geq 0}^+$. Each student $s_i \in S$ has a set $A_i \subseteq P$ of acceptable projects that they rank in order of preference. Ties are allowed in preference lists, where a tie $t$ in a student $s_i$’s list indicates that $s_i$ is indifferent between all projects in $t$. Each lecturer $l_k \in L$ has a preference list over the students $s_i$ for which $A_i \cap P_k \neq \emptyset$. Ties may also exist in lecturer preference lists. The rank of project $p_j$ on student $s_i$’s list, denoted $\text{rank}(s_i, p_j)$, is defined as 1 plus the number of projects that $s_i$ strictly prefers to $p_j$. An analogous definition exists for the rank of a student on a lecturer’s list, denoted $\text{rank}(l_k, s_i)$.

An assignment $M$ in $I$ is a subset of $S \times P$ such that, for each pair $(s_i, p_j) \in M$, $p_j \in A_i$, that is, $s_i$ finds $p_j$ acceptable. Let $M(p_j)$ denote the set of students assigned to a project $p_j \in P$, and let $M(l_k)$ denote the set of students assigned to projects in $P_k$ for a given lecturer $l_k \in L$. A matching $M$ is an assignment such that $|M(s_i)| \leq 1$ for all $s_i \in S$, $|M(p_j)| \leq c_j$ for all $p_j \in P$ and $|M(l_k)| \leq d_k$ for all $l_k \in L$. If $s_i \in S$ is assigned in a matching $M$, we let $M(s_i)$ denote $s_i$’s assigned project, otherwise $M(s_i)$ is empty.

Given a matching $M$ in $I$, let $(s_i, p_j) \in (S \times P) \setminus M$ be a student-project pair, where $p_j$ is offered by lecturer $l_k$. Then $(s_i, p_j)$ is a blocking pair of $M$ if 1, 2 and 3 hold as follows:

1. $s_i$ finds $p_j$ acceptable;
2. $s_i$ either prefers $p_j$ to $M(s_i)$ or is unassigned in $M$;
3. Either $a$, $b$ or $c$ holds as follows:
   a. $p_j$ is undersubscribed (i.e., $|M(p_j)| < c_j$) and $l_k$ is undersubscribed (i.e., $|M(l_k)| < d_k$);
   b. $p_j$ is undersubscribed, $l_k$ is full and either $s_i \in M(l_k)$ or $l_k$ prefers $s_i$ to the worst student in $M(l_k)$;
   c. $p_j$ is full and $l_k$ prefers $s_i$ to the worst student in $M(p_j)$.

Let $(s_i, p_j)$ be a blocking pair of $M$. Then we say that $(s_i, p_j)$ is of type $(3x)$ if 1, 2 and 3x are true in the above definition, where $x \in \{a, b, c\}$. In order to more easily describe certain stages of the approximation algorithm, blocking pairs of type $(3b)$ are split into two subtypes as follows. (3bi) defines a blocking pair of type $(3b)$ where $s_i$ is already assigned to another project of $l_k$’s. (3bii) defines a blocking pair of type $(3b)$ where this is not the case.

A matching $M$ in an instance $I$ of SPA-ST is stable if it admits no blocking pair. Define $\text{MAX SPA-ST}$ to be the problem of finding a maximum stable matching in SPA-ST and let $M_{\text{opt}}$ denote a maximum stable matching for a given instance. Similarly, let $\text{MIN SPA-ST}$ be the problem of finding a minimum stable matching in SPA-ST.

3 Approximation algorithm

3.1 Introduction and preliminary definitions

We begin by defining key terminology before describing the approximation algorithm itself in Section 3.2 which is a non-trivial extension of Király’s HRT algorithm [6].

A student $s_i \in S$ is either in phase 1, 2 or 3. In phase 1 there are still projects on $s_i$’s list that they have not applied to. In phase 2, $s_i$ has iterated once through their list and are doing so again whilst a priority is given to $s_i$ on each lecturer’s preference list, compared to
other students who tie with $s_i$. In phase 3, $s_i$ is considered unassigned and carries out no more applications. A project $p_j$ is fully available if $p_j$ and $l_k$ are both undersubscribed, where lecturer $l_k$ offers $p_j$. A student $s_i$ meta-prefers project $p_j$ to $p_{j'}$ if either (i) $\text{rank}(s_i, p_{j'}) < \text{rank}(s_i, p_j)$, or (ii) $\text{rank}(s_i, p_j) = \text{rank}(s_i, p_{j'})$ and $p_j$ is fully available, whereas $p_{j'}$ is not.

In phase 1 or 2, $s_i$ may be either available, provisionally assigned or confirmed. Student $s_i$ is available if they are not assigned to any project. Student $s_i$ is provisionally assigned if $s_i$ has been assigned in phase 1 and there is a project still on $s_i$’s list that meta-prefers to $p_j$. Otherwise, $s_i$ is confirmed.

If a student $s_i$ is a provisionally assigned to project $p_j$, then $(s_i, p_j)$ is said to be precarious. A project $p_j$ is precarious if it is assigned a student $s_i$ such that $(s_i, p_j)$ is precarious. A lecturer is precarious if they offer a project $p_j$ that is precarious. Lecturer $l_k$ meta-prefers $s_{i_1}$ to $s_{i_2}$ if either (i) $\text{rank}(l_k, s_{i_1}) < \text{rank}(l_k, s_{i_2})$, or (ii) $\text{rank}(l_k, s_{i_1}) = \text{rank}(l_k, s_{i_2})$ and $s_{i_1}$ is in phase 2, whereas $s_{i_2}$ is not. The favourite projects $F_i$ of a student $s_i$ are defined as the set of projects on $s_i$’s preference list for which there is no other project on $s_i$’s list meta-preferred to any project in $F_i$. A worst assignee of lecturer $l_k$ is defined to be a student in $M(l_k)$ of worst rank, with priority given to phase 1 students over phase 2 students. Similarly, a worst assignee of lecturer $l_k$ in $M(p_j)$ is defined to be a student in $M(p_j)$ of worst rank, prioritising phase 1 over phase 2 students, where $l_k$ offers $p_j$.

We remark that some of the above terms such as favourite and precarious have been defined for the SPA-ST setting by extending the definitions of the corresponding terms as given by Király in the HRT context [6].

### 3.2 Description of the algorithm

Algorithm 1 begins with an empty matching $M$ which will be built up over the course of the algorithm’s execution. All students are initially set to be available and in phase 1. The algorithm proceeds as follows. While there are still available students in phase 1 or 2, choose some such student $s_i$. Student $s_i$ applies to a favourite project $p_j$ at the head of their list, that is, there is no project on $s_i$’s list that $s_i$ meta-prefers to $p_j$. Let $l_k$ be the lecturer who offers $p_j$. We consider the following cases.

- If $p_j$ and $l_k$ are both undersubscribed then $(s_i, p_j)$ is added to $M$. Clearly if $(s_i, p_j)$ were not added to $M$, it would potentially be a blocking pair of type (3a).

- If $p_j$ is undersubscribed, $l_k$ is full and $l_k$ is precarious where precarious pair $(s_{i'}, p_{j'}) \in M$ for some project $p_{j'}$ offered by $l_k$, then we remove $(s_{i'}, p_{j'})$ from $M$ and add pair $(s_i, p_j)$. This notion of precariousness allows us to find a stable matching of sufficient size even when there are ties in student preference lists (there may also be ties in lecturer preference lists). Allowing a pair $(s_{i'}, p_{j'}) \in M$ to be precarious means that we are noting that $s_{i'}$ has other fully available project options in their preference list at equal rank to $p_{j'}$. Hence, if another student applies to $p_{j'}$ when $p_{j'}$ is full, or to a project offered by $l_k$ where $l_k$ is full, we allow this assignment to happen removing $(s_{i'}, p_{j'})$ from $M$, since there is a chance that the size of the resultant matching could be increased.

- If on the other hand $p_j$ is undersubscribed, $l_k$ is full and $l_k$ meta-prefers $s_i$ to a worst assignee $s_{i'}$, where $(s_{i'}, p_{j'}) \in M$ for some project $p_{j'}$ offered by $l_k$, then we remove $(s_{i'}, p_{j'})$ from $M$ and add pair $(s_i, p_j)$. It makes intuitive sense that if $l_k$ is full and gets an offer to an undersubscribed project from a student $s_i$ that they prefer to a worst assigned student $s_{i'}$, then $l_k$ would want to remove $s_{i'}$ from $p_{j'}$ and take on $s_i$ for $p_j$. Student $s_{i'}$ will subsequently remove $p_{j'}$ from their preference list as $l_k$ will not want to assign to them on re-application. This is done via the Remove-pref method (Algorithm 2).
\textbf{Algorithm 1} 3/2-approximation algorithm for SPA-ST

\textbf{Require:} An instance $I$ of SPA-ST

\textbf{Ensure:} Return a stable matching $M$ where $|M| \geq \frac{3}{2}|M_{opt}|$

1: $M \leftarrow \emptyset$
2: all students are initially set to be available and in phase 1
3: while there exists an available student $s_i \in S$ who is in phase 1 or 2 do
4: let $l_k$ be the lecturer who offers $p_j$
5: $s_i$ applies to a favourite project $p_j \in A(s_i)$
6: if $p_j$ is fully available then
7: $M \leftarrow M \cup \{(s_i, p_j)\}$
8: else if $p_j$ is undersubscribed, $l_k$ is full and $(l_k$ is precarious or $l_k$ meta-prefers $s_i$ to a worst assignee) then \(\triangleright\) according to the worst assignee definition in Section 3.1
9: if $l_k$ is precarious then
10: let $p_{j'}$ be a project in $P_k$ such that there exists $(s_{i'}, p_{j'}) \in M$ that is precarious
11: else \(\triangleright \) $l_k$ is not precarious
12: let $s_{i'}$ be a worst assignee of $l_k$ such that $l_k$ meta-prefers $s_i$ to $s_{i'}$ and let
13: $p_{j'} = M(s_{i'})$
14: Remove-Pref($s_{i'}, p_{j'}$)
15: end if
16: $M \leftarrow M \setminus \{(s_{i'}, p_{j'})\}$
17: if $p_j$ is full and $(p_j$ is precarious or $l_k$ meta-prefers $s_i$ to a worst assignee in $M(p_j))$ then
18: if $p_j$ is precarious then
19: identify a student $s_{i'} \in M(p_j)$ such that $(s_{i'}, p_j)$ is precarious
20: else \(\triangleright \) $p_j$ is not precarious
21: let $s_{i'}$ be a worst assignee of $l_k$ in $M(p_j)$ such that $l_k$ meta-prefers $s_i$ to $s_{i'}$
22: Remove-Pref($s_{i'}, p_j$)
23: end if
24: $M \leftarrow M \setminus \{(s_{i'}, p_{j'})\}$
25: $M \leftarrow M \cup \{(s_i, p_j)\}$
26: else
27: Remove-Pref($s_i, p_j$)
28: end if
29: end while
30: Promote-students($M$)
31: return $M$;

- If $p_j$ is full and precarious then pair $(s_i, p_j)$ is added to $M$ while precarious pair $(s_{i'}, p_j)$ is removed. As before, this allows $s_{i'}$ to potentially assign to other fully available projects at the same rank as $p_j$ on their list. Since $s_{i'}$ does not remove $p_j$ from their preference list, $s_{i'}$ will get another chance to assign to $p_j$ if these other applications to fully available projects at the same rank are not successful.

- If $p_j$ is full and $l_k$ meta-prefers $s_i$ to a worst assignee $s_{i'}$ in $M(p_j)$, then pair $(s_i, p_j)$ is added to $M$ while $(s_{i'}, p_j)$ is removed. As this lecturer’s project is full (and not precarious) the only time they will want to add a student $s_i$ to this project (meaning the removal of another student) is if $s_i$ is preferred to a worst student $s_{i'}$ assigned to that project.
Algorithm 2 \text{Remove-Pref}(s_i, p_j) – remove a project from a student’s preference list

\textbf{Require:} An instance \(I\) of SPA-ST and a student \(s_i\) and project \(p_j\)

\textbf{Ensure:} Return an instance \(I\) where \(p_j\) is removed from \(s_i\)’s preference list

1: remove \(p_j\) from \(s_i\)’s preference list
2: if \(s_i\)’s preference list is empty then
   3: reinstate \(s_i\)’s preference list
   4: if \(s_i\) is in phase 1 then
      5: move \(s_i\) to phase 2
   6: else if \(s_i\) is in phase 2 then
      7: move \(s_i\) to phase 3
8: end if
9: end if
10: return \(I\)

Algorithm 3 \text{Promote-students}(M) – remove all blocking pairs of type \((3bi)\)

\textbf{Require:} SPA-ST instance \(I\) and matching \(M\) that does not contain blocking pairs of type \((3a)\), \((3bi)\) or \((3c)\).

\textbf{Ensure:} Return a stable matching \(M\).

1: while there are still blocking pairs of type \((3bi)\) do
2: \(s_i, p_{j'}\) be a blocking pair of type \((3bi)\)
3: \(M \leftarrow M \setminus \{(s_i, M(s_i))\}\)
4: \(M \leftarrow M \cup \{(s_i, p_{j'})\}\)
5: end while
6: return \(M\)

Similar to before, \(s_{j'}\) will not subsequently be able to assign to this project and so removes it from their preference list via the Remove-pref method (Algorithm 2).

When removing a project from a student \(s_i\)’s preference list (the Remove-pref operation of Algorithm 2), if \(s_i\) has removed all projects from their preference list and is in phase 1 then their preference list is reinstated and they are set to be in phase 2. If on the other hand they were already in phase 2, then they are set to be in phase 3 and are hence inactive. The proof that Algorithm 2 produces a stable matching (see Appendix B) relies only on the fact that a student iterates once through their preference list. Allowing students to iterate through their preference lists a second time when in phase 2 allows us to find a stable matching of sufficient size when there are ties in lecturer preference lists (there may also be ties in student preference lists). This is due to the meta-prefers definition where a lecturer favours one student \(s_i\) over another \(s_{i'}\) if they are the same rank and \(s_i\) is in phase 2 whereas \(s_{i'}\) is not. Similar to above, this then allows \(s_i\) to steal a position from \(s_{i'}\) with the chance that \(s_{i'}\) may find another assignment and increase the size of the resultant matching.

After the main while loop has terminated, the final part of the algorithm begins where all blocking pairs of type \((3bi)\) are removed using the Promote-students method (Algorithm 3).

3.3 Proof of correctness

\textbf{Theorem 1.} Let \(M\) be a matching found by Algorithm 2 for an instance \(I\) of SPA-ST. Then \(M\) is stable and \(|M| \geq \frac{3}{2} |M_{opt}|\), where \(M_{opt}\) is a maximum stable matching in \(I\).

\textbf{Proof.} Theorems 18, 22 and Theorem 30 proved in Appendix B show that \(M\) is stable, and
that Algorithm 1 runs in polynomial time and has performance guarantee $\frac{2}{3}$. The proofs required for this algorithm are naturally longer and more complex than given by Király [6] for SMTI as SPA-ST generalises SMTI to the case that lecturers can offer multiple projects, and projects and lecturers may have capacities greater than 1. These extensions add extra components to the definition of a blocking pair (given in Section 2) which in turn adds complexity to the algorithm and its proof of correctness.

Appendix B.5 shows a simple example instance where a matching found by Algorithm 1 is exactly $\frac{2}{3}$ times the optimal size, hence the analysis of the performance guarantee is tight.

## 4 IP model

In this section we present an IP model for MAX SPA-ST. For the stability constraints in the model, it is advantageous to use an equivalent condition for stability, as given by the following lemma, whose proof can be found in Appendix C.

▶ Lemma 2. Let $I$ be an instance of SPA-ST and let $M$ be a matching in $I$. Then $M$ is stable if and only if the following condition, referred to as condition (*) holds: For each student $s_i \in S$ and project $p_j \in P$, if $s_i$ is unassigned in $M$ and finds $p_j$ acceptable, or $s_i$ prefers $p_j$ to $M(s_i)$, then either:

- $k$ is full, $s_i \notin M(k)$ and $k$ prefers the worst student in $M(k)$ to $s_i$ or is indifferent between them, or;
- $p_j$ is full and $k$ prefers the worst student in $M(p_j)$ to $s_i$ or is indifferent between them, where $k$ is the lecturer offering $p_j$.

The key variables in the model are binary-valued variables $x_{ij}$, defined for each $s_i \in S$ and $p_j \in P$, where $x_{ij} = 1$ if and only if student $s_i$ is assigned to project $p_j$. Additionally, we have binary-valued variables $\alpha_{ij}$ and $\beta_{ij}$ for each $s_i \in S$ and $p_j \in P$. These variables allow us to more easily describe the stability constraints below. For each $s_i \in S$ and $l_k \in L$, let

$$T_{ik} = \{s_u \in S : \text{rank}(l_k, s_u) \leq \text{rank}(l_k, s_i) \land s_u \neq s_i\}.$$

That is, $T_{ik}$ is the set of students ranked at least as highly as student $s_i$ in lecturer $l_k$’s preference list not including $s_i$. Also, for each $p_j \in P$, let

$$T_{ijk} = \{s_u \in S : \text{rank}(l_k, s_u) \leq \text{rank}(l_k, s_i) \land s_u \neq s_i \land p_j \in A(s_u)\}.$$

That is, $T_{ijk}$ is the set of students $s_u$ ranked at least as highly as student $s_i$ in lecturer $l_k$’s preference list, such that project $p_j$ is acceptable to $s_u$, not including $s_i$. Finally, let $S_{ij} = \{p_r \in P : \text{rank}(s_i, p_r) \leq \text{rank}(s_i, p_j)\}$, that is, $S_{ij}$ is the set of projects ranked at least as highly as project $p_j$ in student $s_i$’s preference list, including $p_j$. Figure 1 shows the IP model for MAX SPA-ST.

Equation (1) enforces $x_{ij} = 0$ if $s_i$ finds $p_j$ unacceptable. Inequality (2) ensures that a student may be assigned to a maximum of one project. Inequalities (3) and (4) ensure that project and lecturer capacities are enforced. In the left hand side of Inequality (5), if $1 - \sum_{p_r \in S_{ij}} x_{ir} = 1$, then either $s_i$ is unmatched or $s_i$ prefers $p_j$ to $M(s_i)$. This also ensures that either $\alpha_{ij} = 1$ or $\beta_{ij} = 1$, described in Inequalities (6) and (7). Inequality (6) ensures that, if $\alpha_{ij} = 1$, the number of students ranked at least as highly as student $s_i$ by $l_k$ (not including $s_i$) and assigned to $l_k$ must be at least $l_k$’s capacity $d_k$. Inequality (7) ensures that, if $\beta_{ij} = 1$, the number of students ranked at least as highly as student $s_i$ in lecturer $l_k$’s preference list (not including $s_i$) and assigned to $p_j$ must be at least $p_j$’s capacity $c_j$. 
A \( \frac{3}{2} \)-approximation algorithm for the Student-Project Allocation problem

maximise: \[ \sum_{s_i \in S} \sum_{p_j \in P} x_{ij} \]

subject to:
1. \( x_{ij} = 0 \) \quad \forall s_i \in S \forall p_j \in P, p_j \notin A(s_i) \)
2. \( \sum_{p_j \in P} x_{ij} \leq 1 \) \quad \forall s_i \in S \forall p_j \in P
3. \( \sum_{s_i \in S} x_{ij} \leq c_j \) \quad \forall p_j \in P \forall l_k \in L
4. \( \sum_{s_i \in S} \sum_{p_j \in P} x_{ij} \leq d_k \) \quad \forall l_k \in L
5. \( 1 - \sum_{p_r \in S_{ij}} x_{ir} \leq \alpha_{ij} + \beta_{ij} \) \quad \forall s_i \in S \forall p_j \in P
6. \( \sum_{s_u \in T_{ik}} \sum_{p_r \in P_k} x_{ur} \geq d_k \alpha_{ij} \) \quad \forall s_i \in S \forall p_j \in P
7. \( \sum_{s_u \in T_{ik}} x_{uj} \geq c_j \beta_{ij} \) \quad \forall s_i \in S \forall p_j \in P

\( x_{ij} \in \{0, 1\}, \quad \alpha_{ij} \in \{0, 1\}, \quad \beta_{ij} \in \{0, 1\} \) \quad \forall s_i \in S \forall p_j \in P

\textbf{Figure 1} IP model for \textbf{MAX SPA-ST}

Finally, for our optimisation we maximise the sum of all \( x_{ij} \) variables in order to maximise the number of students assigned. The following result, proved in Appendix C, establishes the correctness of the IP model.

\textbf{Theorem 3.} Given an instance \( I \) of \textbf{SPA-ST}, let \( J \) be the IP model as defined in Figure 1. A maximum stable matching in \( I \) corresponds to an optimal solution in \( J \) and vice versa.

\section{Experimental evaluation}

\subsection{Methodology}

Experiments were conducted on the approximation algorithm and the IP model using randomly-generated data in order to measure the effects on matching statistics when changing parameter values relating to (1) instance size, (2) probability of ties in preference lists, and (3) preference list lengths. Two further experiments (referred to as (4) and (5) below) explored scalability properties for both techniques. Instances were generated using both existing and new software. The existing software is known as the \textit{Matching Algorithm Toolkit} and is a collaborative project developed by students and staff at the University of Glasgow.

For a given \textbf{SPA-ST} instance, let the total project and lecturer capacities be denoted by \( c_P \) and \( d_L \), respectively. Note that these capacities were distributed randomly, subject to there being a maximum difference of 1 between the capacities of any two projects or any two lecturers (to ensure uniformity). The minimum and maximum size of student preference lists is given by \( l_{\text{min}} \) and \( l_{\text{max}} \), and \( t_s \) represents the probability that a project on a student’s preference list is tied with the next project. Lecturer preference lists were generated initially from the student preference lists, where a lecturer \( l_k \) must rank a student if a student ranks...
Preference list scalability:

Instance size scalability:

Increasing probability of ties:

Increasing instance size:

A linear distribution was used to make some projects more popular than others and in all experiments the most popular project is around 5 times more popular than the least. This distribution influenced the likelihood of a student finding a given project acceptable. Parameter details for each experiment are given below.

(1) Increasing instance size: 10 sets of 10,000 instances were created (labelled SIZE1, ..., SIZE10). The number of students \( n_1 \) increased from 100 to 1000 in steps of 100, with \( n_2 = 0.6n_1, n_3 = 0.4n_1 \), \( c_P = 1.4n_1, d_L = 1.2n_1 \). The probabilities of ties in preference lists were \( t_s = t_l = 0.2 \) throughout all instance sets. Lengths of preference lists \( l_{\text{min}} = 3 \) and \( l_{\text{max}} = 5 \) also remained the same and were kept low to ensure a wide variability in stable matching size per instance.

(2) Increasing probability of ties: 11 sets of 10,000 instances were created (labelled TIES1, ..., TIES11). Throughout all instance sets \( n_1 = 300, n_2 = 250, n_3 = 120, c_P = 420, d_L = 360, l_{\text{min}} = 3 \) and \( l_{\text{max}} = 5 \). The probabilities of ties in student and lecturer preference lists increased from \( t_s = t_l = 0.0 \) to \( t_s = t_l = 0.5 \) in steps of 0.05.

(3) Increasing preference list lengths: 10 sets of 10,000 instances were generated (labelled PREF1, ..., PREF10). Similar to the TIES cases, throughout all instance sets \( n_1 = 300, n_2 = 250, n_3 = 120, c_P = 420 \) and \( d_L = 360 \). Additionally, \( t_s = t_l = 0.2 \). Preference list lengths increased from \( l_{\text{min}} = l_{\text{max}} = 1 \) to \( l_{\text{min}} = l_{\text{max}} = 10 \) in steps of 1.

(4) Instance size scalability: 5 sets of 10 instances were generated (labelled SCALS1, ..., SCALS5). All instance sets in this experiment used the same parameter values as the SIZE experiment, except the number of students \( n_1 \) increased from 10,000 to 50,000 in steps of 10,000.

(5) Preference list scalability: Finally, 6 sets of 10 instances were created (labelled SCALP1, ..., SCALP6). Throughout all instance sets \( n_1 = 500 \) with the same values for other parameters as the SIZE experiment. However in this case ties were fixed at \( t_s = t_l = 0.4 \), and \( l_{\text{min}} = l_{\text{max}} \) increasing from 25 to 150 in steps of 25.

For each generated instance, we ran the \( \frac{2}{3} \)-approximation algorithm and then used the IP model to find a maximum stable matching. We also computed a minimum stable matching using a simple adaptation of our IP model for \( \text{MAX SPA-ST} \) in order to measure the spread in the sizes of stable matchings. A timeout of 1800 seconds (30 minutes) was imposed on all instance runs. All experiments were conducted using a machine with 32 cores, 8×64GB RAM and Dual Intel® Xeon® CPU E5-2697A v4 processors. The operating system was Ubuntu version 17.04 with all code compiled in Java version 1.8, where the IP models were solved using Gurobi version 7.5.2. Each approximation algorithm instance was run on a single thread while each IP instance was run on two threads. No attempt was made to parallelise Java garbage collection. Repositories for the code and data can be found at https://doi.org/10.5281/zenodo.1183221 and https://doi.org/10.5281/zenodo.1186823 respectively.

Correctness testing was conducted over all generated instances. This consisted of (1) ensuring that each matching produced by the approximation algorithm was at least \( \frac{2}{3} \) the size of maximum stable matching, as found by the IP and, (2) testing that a given allocation was stable and adhered to all project and lecturer capacities. This was run over all output from both the approximation algorithm and the IP-based algorithm.

5.2 Experimental results

Experimental results can be seen in Tables 1, 2, 3 and 4. Tables 1, 2 and 3 show the results from Experiments 1, 2 and 3 respectively (in which the instance size, probability of
ties and preference list lengths were increased, respectively). From this point onwards an optimal matching refers to a maximum stable matching. In these tables, column ‘minimum A/Max’ gives the minimum ratio of approximation algorithm matching size to optimal matching size that occurred, ‘% A=Max’ displays the percentage of times the approximation algorithm achieved an optimal result, and ‘% A ≥ 0.98 A/Max’ shows the percentage of times the approximation algorithm achieved a result at least 98% of optimal. The ‘average size’ columns are somewhat self explanatory, with sub-columns ‘A/Max’ and ‘Min/Max’ showing the average approximation algorithm matching size and minimum stable matching size as a fraction of optimal. Finally, ‘average total time’ indicates the time taken for model creation, solving and outputting results per instance. The main findings are summarised below.

- The approximation algorithm consistently far exceeds its $\frac{3}{2}$ bound. Considering the column labelled ‘minimum A/Max’ in Tables 1, 2 and 3, we see that the smallest value was within the SIZE1 instance set with a ratio of 0.9286. This is well above the required bound of $\frac{3}{2}$.

- On average the approximation algorithm provides results that are closer in size to the average maximum stable matching than the minimum stable matching. The columns ‘A/Max’ and ‘Min/Max’ show that, on average, for each instance set, the approximation algorithm produces a solution that is within 98% of maximum and far closer to the maximum size than to the minimum size.

Table 4 shows the scalability results for increasing instance sizes (Experiment 4) and increasing preference list lengths (Experiment 5). The ‘instances completed’ column indicates the number of instances completed before timeout occurred. In addition to showing the average total time taken (where ‘total’ includes model creation time and solution time), the column ‘average solve time’ displays the time taken to either execute the approximation algorithm, or solve the IP model (in both cases, model creation time is excluded).

For Experiment 4, the number of instances solved within the 30-minute timeout reduced from 10 to 0 for the IP-based algorithm finding the maximum stable matching. However, even for the largest instance set sizes the approximation algorithm was able to solve all instances on average within a total of 21 seconds (0.8 seconds of which was used to actually execute the algorithm).

For Experiment 5, with a higher probability of ties and increasing preference list lengths, the IP-based algorithm was only able to solve all the instances of one instance set (SCALP2) within 30 minutes each, however the approximation algorithm took less than 0.3 seconds on average to return a solution for each instance. This shows that the approximation algorithm is useful for either larger or more complex instances than the IP-based algorithm can handle, motivating its use for real world scenarios.

6 Future work

This paper has described a $\frac{3}{2}$-approximation algorithm for MAX SPA-ST. It remains open to describe an approximation algorithm that has a better performance guarantee, and/or to prove a stronger lower bound on the inapproximability of the problem than the current best bound of $\frac{33}{29}$ [12]. Further experiments could also measure the extent to which the order that students apply to projects in Algorithm 1 affects the size of the stable matching generated.

The work in this paper has mainly focused on the size of stable matchings. However, it is possible for a stable matching to admit a blocking coalition, where a permutation of student assignments could improve the allocations of the students and lecturers involved without harming anyone else. Since permutations of this kind cannot change the size of the matching they are not studied further here, but would be of interest for future work.
### Table 1: Increasing instance size experimental results.

| Case   | minimum | A=Max | % A≥ 0.98Max | A | Min | average size | average total time (ms) |
|--------|---------|-------|--------------|---|-----|--------------|------------------------|
| SIZE1  | 0.9286  | 17.8  | 62.7         | 96.4 | 92.0 | 97.8 | 0.986 | 0.941 | 43.3 | 147.6 | 137.8 |
| SIZE2  | 0.9585  | 1.6   | 62.6         | 192.6 | 183.4 | 195.7 | 0.984 | 0.937 | 51.2 | 230.6 | 210.6 |
| SIZE3  | 0.9556  | 0.1   | 63.7         | 288.7 | 274.9 | 293.7 | 0.983 | 0.936 | 56.6 | 346.4 | 313.4 |
| SIZE4  | 0.9644  | 0.0   | 65.6         | 384.9 | 366.4 | 391.7 | 0.983 | 0.935 | 59.7 | 488.7 | 429.3 |
| SIZE5  | 0.9654  | 0.0   | 66.5         | 481.0 | 457.7 | 489.6 | 0.982 | 0.935 | 62.8 | 660.3 | 555.6 |
| SIZE6  | 0.9641  | 0.0   | 66.8         | 577.2 | 549.3 | 587.7 | 0.982 | 0.935 | 66.4 | 862.3 | 713.0 |
| SIZE7  | 0.9679  | 0.0   | 65.4         | 673.3 | 640.5 | 685.7 | 0.982 | 0.934 | 69.8 | 1127.8 | 900.6 |
| SIZE8  | 0.9684  | 0.0   | 67.4         | 769.5 | 732.0 | 783.8 | 0.982 | 0.934 | 73.0 | 1437.3 | 1098.2 |
| SIZE9  | 0.9653  | 0.0   | 68.6         | 865.6 | 823.4 | 881.7 | 0.982 | 0.934 | 76.5 | 1784.3 | 1343.9 |
| SIZE10 | 0.9701  | 0.0   | 68.0         | 961.7 | 914.7 | 979.7 | 0.982 | 0.934 | 86.6 | 2281.2 | 1651.0 |

### Table 2: Increasing probability of ties experimental results.

| Case   | minimum | A=Max | % A≥ 0.98Max | A | Min | average size | average total time (ms) |
|--------|---------|-------|--------------|---|-----|--------------|------------------------|
| TIES1  | 1.0000  | 100.0 | 100.0        | 284.0 | 284.0 | 284.0 | 1.000 | 1.000 | 59.2 | 184.0 | 186.9 |
| TIES2  | 0.9792  | 38.0  | 100.0        | 284.9 | 282.0 | 285.8 | 0.997 | 0.987 | 61.2 | 192.4 | 194.7 |
| TIES3  | 0.9722  | 12.1  | 99.3         | 285.9 | 279.9 | 287.9 | 0.993 | 0.972 | 61.7 | 201.0 | 203.1 |
| TIES4  | 0.9655  | 3.4   | 95.2         | 287.0 | 277.6 | 289.9 | 0.990 | 0.958 | 62.3 | 213.3 | 214.5 |
| TIES5  | 0.9626  | 1.0   | 82.5         | 288.0 | 275.1 | 291.9 | 0.986 | 0.942 | 62.9 | 234.3 | 231.0 |
| TIES6  | 0.9558  | 0.4   | 66.7         | 289.2 | 272.4 | 294.0 | 0.984 | 0.927 | 64.2 | 274.2 | 260.6 |
| TIES7  | 0.9486  | 0.2   | 52.9         | 290.3 | 269.4 | 295.7 | 0.982 | 0.911 | 64.3 | 358.3 | 311.3 |
| TIES8  | 0.9527  | 0.2   | 46.4         | 291.4 | 266.2 | 297.2 | 0.980 | 0.896 | 64.2 | 577.3 | 380.7 |
| TIES9  | 0.9467  | 0.2   | 50.4         | 292.5 | 262.7 | 298.3 | 0.980 | 0.880 | 65.2 | 1234.1 | 427.5 |
| TIES10 | 0.9529  | 0.5   | 61.9         | 293.7 | 258.9 | 299.1 | 0.982 | 0.866 | 59.6 | 2903.4 | 409.1 |
| TIES11 | 0.9467  | 1.0   | 74.2         | 294.8 | 254.8 | 299.5 | 0.984 | 0.851 | 60.4 | 5756.9 | 377.4 |
8:12

A \approx\text{-}approximation algorithm for the Student-Project Allocation problem

\[ A = \text{Max} \geq 0.98 \text{Max} \]

\[ \text{average size} \]

\[ \text{average total time (ms)} \]

\begin{tabular}{lccccccccc}
\hline
Case & minimum A/Max & % A=Max & % A \geq 0.98Max & average A & Min & Max & A/Max & Min/Max & A & Min & Max \\
\hline
PREF1 & 1.0000 & 100.0 & 100.0 & 215.0 & 215.0 & 215.0 & 1.000 & 1.000 & 74.3 & 107.5 & 105.1 \\
PREF2 & 0.9699 & 12.3 & 99.0 & 262.1 & 249.1 & 264.1 & 0.993 & 0.943 & 67.5 & 133.8 & 128.7 \\
PREF3 & 0.9617 & 12.3 & 84.0 & 280.9 & 266.4 & 284.7 & 0.987 & 0.936 & 68.1 & 181.4 & 174.0 \\
PREF4 & 0.9623 & 1.0 & 82.8 & 290.0 & 277.0 & 293.9 & 0.987 & 0.943 & 69.1 & 249.7 & 242.6 \\
PREF5 & 0.9661 & 4.2 & 95.1 & 294.8 & 283.9 & 297.7 & 0.990 & 0.954 & 68.3 & 346.7 & 340.3 \\
PREF6 & 0.9732 & 15.7 & 99.5 & 297.3 & 288.7 & 299.1 & 0.994 & 0.965 & 66.1 & 472.4 & 440.6 \\
PREF7 & 0.9767 & 36.2 & 100.0 & 298.7 & 292.1 & 299.7 & 0.997 & 0.975 & 64.5 & 638.3 & 550.9 \\
PREF8 & 0.9833 & 58.2 & 100.0 & 299.3 & 294.4 & 299.9 & 0.998 & 0.982 & 64.1 & 811.9 & 660.3 \\
PREF9 & 0.9866 & 75.5 & 100.0 & 299.7 & 296.1 & 299.9 & 0.999 & 0.987 & 63.4 & 1032.2 & 789.1 \\
PREF10 & 0.9900 & 87.3 & 100.0 & 299.8 & 297.4 & 300.0 & 1.000 & 0.991 & 104.3 & 1239.4 & 931.0 \\
\hline
\end{tabular}

Table 3: Increasing preference list length experimental results.

\begin{tabular}{lcccccccc}
\hline
Case & instances completed & average solve time (ms) & average total time (ms) & & & & & & & \\
& A & Min & Max & A & Min & Max & A & Min & Max \\
\hline
SCALS1 & 10 & 10 & 10 & 136.5 & 126162.8 & 225917.9 & 1393.8 & 127980.3 & 227764.3 \\
SCALS2 & 10 & 10 & 9 & 242.4 & 348849.4 & 1091424.2 & 5356.7 & 353272.3 & 1096045.6 \\
SCALS3 & 10 & 10 & 0 & 491.7 & 777267.7 & N/A & 13095.3 & 785421.2 & N/A \\
SCALS4 & 10 & 7 & 0 & 718.8 & 1049122.0 & N/A & 18883.5 & 1062076.4 & N/A \\
SCALS5 & 10 & 7 & 0 & 803.5 & 1288961.1 & N/A & 20993.0 & 1307728.7 & N/A \\
SCALP1 & 10 & 0 & 9 & 25.1 & N/A & 93086.0 & 193.3 & N/A & 94242.9 \\
SCALP2 & 10 & 1 & 10 & 23.3 & 1425177.0 & 626774.9 & 189.4 & 1428844.0 & 631225.2 \\
SCALP3 & 10 & 0 & 3 & 31.7 & N/A & 867107.7 & 196.6 & N/A & 882251.0 \\
SCALP4 & 10 & 0 & 1 & 37.8 & N/A & 1551376.0 & 248.5 & N/A & 1594201.0 \\
SCALP5 & 10 & 0 & 0 & 59.0 & N/A & N/A & 283.7 & N/A & N/A \\
SCALP6 & 10 & 0 & 0 & 45.7 & N/A & N/A & 288.4 & N/A & N/A \\
\hline
\end{tabular}

Table 4: Scalability experimental results.
References

1. D.J. Abraham, K. Cechlárová, D.F. Manlove, and Kurt Melhorn. Pareto optimality in house allocation problems. In Proceedings. ISAAC ’04: The 15th Annual International Symposium on Algorithms and Computation, Lecture Notes in Computer Science, volume 3341, pages 3 – 15. Springer, 2004.
2. D.J. Abraham, R.W. Irving, and D.F. Manlove. Two algorithms for the student-project allocation problem. Journal of Discrete Algorithms, 5:73–90, 2007.
3. G. Brassard and P. Bratley. Fundamentals of Algorithmics. Prentice-Hall, 1996.
4. R. Calvo-Serrano, G. Guillén-Gosálbez, S. Kohn, and A. Masters. Mathematical programming approach for optimally allocating students’ projects to academics in large cohorts. Education for Chemical Engineers, 20:11–21, 2017.
5. M. Chiarandini, R. Fagerberg, and S. Gualandi. Handling preferences in student-project allocation. Annals of Operations Research, to appear, 2018.
6. Z. Király. Linear time local approximation for maximum stable marriage. Algorithms, 6:471–484, 2013.
7. A. Kwanashie, R.W. Irving, D.F. Manlove, and C.T.S. Sng. Profile-based optimal matchings in the Student–Project Allocation problem. In Proceedings of IWOCA ’14: the 25th International Workshop on Combinatorial Algorithms, volume 8986 of Lecture Notes in Computer Science, pages 213–225. Springer, 2015.
8. D.F. Manlove. Algorithmics of Matching Under Preferences. World Scientific, 2013.
9. D.F. Manlove, R.W. Irving, K. Iwama, S. Miyazaki, and Y. Morita. Hard variants of stable marriage. Theoretical Computer Science, 276(1-2):261–279, 2002.
10. D.F. Manlove and G. O’Malley. Student-project allocation with preferences over projects. Journal of Discrete Algorithms, 6:553–560, 2008.
11. A.E. Roth. The evolution of the labor market for medical interns and residents: a case study in game theory. Journal of Political Economy, 92(6):991–1016, 1984.
12. H. Yanagisawa. Approximation Algorithms for Stable Marriage Problems. PhD thesis, Kyoto University, School of Informatics, 2007.

Appendix

A Cloning from SPA-ST to HRT

Manlove [8, Theorem 3.11] describes a polynomial transformation from a weakly stable matching in and instance of HRT to a weakly stable matching in an instance of SMIT, and vice versa, where the size of matchings is conserved.

A natural cloning method to convert instances of SPA-ST to instances of HRT is given as Algorithm 4. This algorithm involves converting students into residents and projects into hospitals. Hospitals inherit their capacity from projects. Residents inherit their preference lists naturally from students. Hospitals inherit their preference lists from the lecturer who offers their associated project; a resident entry $r_i$ is ranked only if $r_i$ also ranks this hospital. In order to translate lecturer capacities into the HRT instance, a number of dummy residents are created for each lecturer. The number created for lecturer $l_k$ is equal to the sum of capacities of their offered projects $P_k$ minus the capacity of $l_k$. We will ensure that all dummy residents are assigned in any stable matching. To this end, each dummy resident has a first position tie of all hospitals associated with projects of $l_k$, and each hospital $h_j$ in this set has a first position tie of all dummy residents associated with $l_k$. In this way, as all dummy
residents must be assigned in any stable matching by Proposition 4, lecturer capacities are automatically adhered to.

**Proposition 4.** Let $I'$ be an instance of HRT created from an instance $I$ of SPA-ST using Algorithm 4. All dummy residents must be assigned in any stable matching in $I'$.

**Proof.** In $I'$, for each lecturer $l_k$, the number of dummy residents created is equal to $f_k = \sum_{p_j \in P_k} c_j - d_k$. Assume for contradiction that one of the dummy residents $r_{d_1}^k$ is unassigned in some stable matching $M'$ of $I'$.

Let $H_k$ denote the set of hospitals associated with projects of $l_k$. Since $r_{d_1}^k$ is a dummy resident, it must have all hospitals in $H_k$ tied in first position. Also, each hospital in $H_k$ must rank all $f_k$ dummy students (associated with $l_k$) in tied first position. Since $r_{d_1}^k$ is unassigned in $M'$, $r_{d_1}^k$ would prefer to be assigned to any hospital in $H_k$. Also, since there is at least one dummy resident unassigned, there must be at least one hospital $h_{d_2}^k$ in $H_k$ that has fewer first-choice assignees than its capacity. Hospital $h_{d_2}^k$ must exist since if it did not, then all dummy residents would be matched. But then $(r_{d_1}^k, h_{d_2}^k)$ would be a blocking pair of $M'$, a contradiction. ▶

**Algorithm 4** Clone-SPA-ST. Converts an SPA-ST instance into an HRT instance

**Require:** An instance $I$ of SPA-ST

**Ensure:** Return an instance $I'$ of HRT

1. for all student $s_i$ in $S$ do
   2. create a resident $r_i$
   3. $r_i$ inherits their preference list from $s_i$’s list, ranking hospitals rather than projects
4. end for
5. for all project $p_j$ in $P$ do
   6. create a hospital $h_j$
   7. $h_j$’s capacity is given by $e_j = c_j$
   8. let $l_k$ be the lecturer offering project $p_j$
   9. $h_j$ inherits their preference list from $l_k$’s list, where a resident entry $r_i$ is retained only if $r_i$ also ranks $h_j$
10. end for
11. for all lecturer $l_k$ in $L$ do
12.   if $d_k < \sum_{p_j \in P_k} c_j$ then
13.     let $f_k = \sum_{p_j \in P_k} c_j - d_k$
14.     create $f_k$ new dummy residents $r^k = \{r_1^k, r_2^k, \ldots, r_f^k\}$
15.     let $H_k$ denote the set of all hospitals in $I'$ associated with the projects of $P_k$ in $I$
16.     the preference list of each dummy resident is given by a first position tie of all hospitals in $H_k$
17.     a first position tie of all residents in $r^k$ is added to the start of the preference list of each hospital in $H_k$
18.   end if
19. end for
20. Let HRT instance $I'$ be formed from all residents (including dummy residents) and hospitals
21. return instance $I'$

**Theorem 5.** Given an instance $I$ of SPA-ST we can construct an instance $I'$ of HRT in $O(n_1 + Dn_2 + m)$ time with the property that a stable matching $M$ in $I$ can be converted to a
stable matching $M'$ in $\text{HRT}$ in $O(Dn_2 + m)$ time, where $|M'| = |M| + \sum_{l_k \in Q} \sum_{p_r \in P_k} (c_r) - d_k$. Here, $n_1$ denotes the number of students, $n_2$ the number of projects, $D$ the total capacities of lecturers and $m$ the total length of student preference lists.

**Proof.** Suppose $M$ is a stable matching in $I$. We construct an instance $I'$ of $\text{HRT}$ using Algorithm 4. The time complexity of $O(n_1 + Dn_2 + m)$ for the reduction carried out by the algorithm is achieved by noting that $I'$ has a maximum of $n_1 + n_2 + D$ agents and that there are a maximum of $Dn_2 + m$ acceptable resident-hospital pairs.

Initially let $M' = M$. Let the set of dummy residents $R_d$ in $I'$ form a resident-complete matching (when considering the residents in $R_d$) with the set of all hospitals and add these pairs to $M'$. This is possible (as proved in Proposition 4), because for each lecturer $l_k$ each dummy resident associated with $l_k$ denoted $R_d^k$ finds all projects in $P_k$ acceptable, and moreover the total number of remaining positions of the projects is at least as large as $\sum_{p_r \in P_k} (c_r) - |M(l_k)| = \delta_k$ and $|R_d^k| = \delta_k$ by definition.

We claim that $M'$ is stable in $I'$. Suppose for contradiction that $(r_i, h_j)$ blocks $M'$ in $I'$.

- All dummy residents must be assigned in $M'$ to their first-choice hospital by above, hence $r_i$ corresponds to a student $s_i$ in $I$. Resident $r_i$ inherited their preference list from $s_i$ hence we know that $s_i$ finds $p_j$ acceptable. Therefore by the definition given in Section 2 condition (1) of a blocking pair of $M$ in $I$ is satisfied.
- Resident $r_i$ is either unassigned in $M'$ or prefers $h_j$ to $M'(r_i)$. Student $s_i$ is therefore in an equivalent position and condition (2) of a blocking pair of $M$ in $I$ is satisfied.
- Hospital $h_j$ is either undersubscribed or prefers $r_i$ to their worst assignee in $M'$.
  - If $h_j$ is undersubscribed, then $p_j$ must also be undersubscribed. If $l_k$ were full in $M$, then $h_j$ would be full in $M'$ since $\sum_{p_r \in P_k} (c_r) = \delta_k + |M(l_k)|$ in this scenario, therefore $l_k$ must be undersubscribed. But if $l_k$ is undersubscribed then this satisfies condition (3a) of a blocking pair.
  - If $h_j$ prefers $r_i$ to their worst assignee in $M'$, then $l_k$ must prefer $s_i$ to their worst assignee in $M(p_j)$. This satisfies condition 3(c) of a blocking pair.

Therefore by the definition in Section 2 $(s_i, p_j)$ is a blocking pair of $M$ in $I$, a contradiction.

Since dummy residents are added in the algorithm’s execution it is clear that in general $|M| \neq |M'|$. However, since all dummy residents must be assigned by Proposition 4 it is trivial to calculate the difference

$$|M'| = |M| + |R_d| = |M| + \sum_{l_k \in Q} \sum_{p_r \in P_k} (c_r) - d_k.$$

The converse of Theorem 5 is not true in general, as shown in the example in Figure 2. Here a stable matching $M'$ in an instance $I'$ of $\text{HRT}$ does not convert into a stable matching $M$ of the associated instance $I$ of $\text{SPA-ST}$.

A natural question arises as to whether in special circumstances we can use the cloning process described in Algorithm 4 that is, convert to $\text{HRT}$ then to $\text{SMT}$, then use Király’s 3/2-approximation algorithm on the $\text{SMT}$ instance, and obtain as a result a matching that is a 3/2-approximation to a maximum stable matching in the original instance of $\text{SPA-ST}$. For example, assuming that the matching $M$ resulting from the above process is stable, can we ensure the 3/2 bound?

The following example demonstrates Algorithm 4 in use and shows that the process described above is not sufficient to retain the 3/2-approximation in an $\text{SPA-ST}$ instance even if $M$ is stable.
Table 6. On line m in a stable matching. Therefore allowing the beginning of ability to take instance the same two-stage process as in Figure 3. Finally, Table 6 shows the algorithm trace for instance to an spa-st algorithm to the maximum stable matching specifically for instances of matching for instances of above and Király’s algorithm does not result in a M by M created from dummy residents in Algorithm 4, I of proposals) yield the matching shown in bold.

Figure 2 Conversion of a stable matching M’ in [HRT] into matching M in [SPA-ST].

Algorithm [HRT] is used to convert the [SPA-ST] instance I” of HRT in Figure 3a to an instance I’ of HRT in Figure 3b which is then itself converted to the instance I”’ of SMTI in Figure 3c using the process described by Manlove [8] Theorem 3.11. In this process men correspond to hospitals in I’ (projects in I) and women correspond to residents in I’ (students in I). Executing Király’s [8] 3/2-approximation algorithm on the SMTI instance I” could (depending on order of proposals) yield the matching

\[ M'' = \{(w_2, m_4), (w_4, m_3), (w_5, m_2), (w_6, m_6), (w_7, m_7)\} \]

in I”. A trace of how this matching is created is given in Table 5. As w_5, w_6 and w_7 were created from dummy residents in Algorithm 4 M” (stable in I”) converts into the matching M_e = \{(s_2, p_4), (s_4, p_3)\} of size 2 in I. But a maximum stable matching in I is of size 4, given by M = \{(s_1, p_3), (s_2, p_1), (s_3, p_3), (s_4, p_2)\}. Therefore using the cloning method described above and Király’s algorithm does not result in a 3/2-approximation to the maximum stable matching for instances of SMTI. This motivates the development of a 3/2-approximation algorithm to the maximum stable matching specifically for instances of SMTI.

An intuitive idea as to how the addition of dummy residents in the conversion of this SPA-ST instance to an SMTI instance stops the retention of the 3/2 bound follows. Figure 4a shows the SPA-ST instance I_2 which is the same as the instance in Figure 3a but with project 1 and 2 capacities reduced to 1. Figure 4b shows I_2 converted into an SMTI instance I_2’ using the same two-stage process as in Figure 3. Finally, Table 6 shows the algorithm trace for instance I_2’ using Király’s 3/2 SMTI algorithm.

The main first difference in the traces can be seen on line 14 of Table 5 and line 11 of Table 6. On line 11 of Table 6 m_2 applies to w_4 as an advantaged man, giving them the ability to take w_4 from m_3. This shows the benefit of having a tie including m_2 and m_3 at the beginning of w_4’s list - either of these men being matched to w_4 would be equally useful in a stable matching. Therefore allowing m_2 to take w_4 from m_3 gives m_3 a chance to get another partner, increasing the size of matching eventually attained. On line 14 of Table 5 m_2 becomes ‘stuck’ on w_5, one of the women derived from a dummy resident. This stops m_2...
| Action                                                                 | m₁  | m₂  | m₃  | m₄  | m₅  | m₆  | m₇  |
|-----------------------------------------------------------------------|-----|-----|-----|-----|-----|-----|-----|
| 1  m₇ applies w₇, accepted                                            | u₇  |     |     |     |     |     |     |
| 2  m₆ applies w₅, accepted                                            |     | u₅  | u₇  |     |     |     |     |
| 3  m₅ applies w₆, accepted                                            | w₆  | w₅  | u₇  |     |     |     |     |
| 4  m₄ applies w₇, rejected, w₇ pref removed by m₄                      | w₄  | w₆  | w₅  | u₇  |     |     |     |
| 5  m₄ applies w₄, accepted                                            |     | w₄  | w₅  | u₇  |     |     |     |
| 6  m₃ applies w₇, rejected, w₇ pref removed by m₃                      | w₄  | w₅  | w₆  | u₇  |     |     |     |
| 7  m₃ applies w₄, accepted, w₄ pref removed by m₄                      | w₄  |     | w₅  | u₇  |     |     |     |
| 8  m₄ applies w₂, accepted                                            | w₄  | w₃  | w₅  | u₇  |     |     |     |
| 9  m₂ applies w₅, rejected, w₅ pref removed by m₂                      | w₄  | w₃  | w₅  |     | u₇  |     |     |
| 10 m₂ applies w₆, rejected, w₆ pref removed by m₂                      | w₄  | w₃  | w₅  |     |     | u₇  |     |
| 11 m₁ applies w₆, rejected, w₆ pref removed by m₂                      | w₄  | w₃  | w₅  | u₇  |     |     |     |
| 12 m₃ applies w₄, rejected, w₄ pref removed by m₂                      | w₄  | w₃  | w₅  |     | u₇  |     |     |
| 13 m₂ advantaged                                                       | w₄  | w₃  | w₅  |     |     |     | u₇  |
| 14 m₂ applies w₅, accepted                                            | u₅  | w₄  | w₃  | w₅  | u₇  |     |     |
| 15 m₆ applies w₆, rejected, w₆ pref removed by m₆                      | u₅  | w₄  | w₃  | w₅  |     | u₇  |     |
| 16 m₆ applies w₄, rejected, w₄ removed by m₆                           | u₅  | w₄  | w₃  | w₅  |     |     | u₇  |
| 17 m₆ applies w₄, rejected, w₄ removed by m₆                           | u₅  | w₄  | w₃  | w₅  | u₇  |     |     |
| 18 m₆ advantaged                                                       | u₅  | w₄  | w₃  | w₅  |     |     |     |
| 19 m₆ applies w₅, rejected, w₅ pref removed by m₆                      | u₅  | w₄  | w₃  | w₅  |     |     | u₇  |
| 20 m₆ applies w₆, accepted                                            | u₅  | w₄  | w₃  | u₇  |     |     |     |
| 21 m₅ applies w₅, rejected, w₅ pref removed by m₅                      | u₅  | w₄  | w₃  |     | u₇  |     |     |
| 22 m₅ applies w₆, rejected, w₆ pref removed by m₅                      | u₅  | w₄  | w₃  | u₇  |     |     |     |
| 23 m₅ applies w₄, rejected, w₄ pref removed by m₅                      | u₅  | w₄  | w₃  |     |     | u₇  |     |
| 24 m₅ advantaged                                                       | u₅  | w₄  | w₃  |     |     |     |     |
| 25 m₅ applies w₅, rejected, w₅ pref removed by m₅                      | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 26 m₅ applies w₆, rejected, w₆ pref removed by m₅                      | u₅  | w₄  | w₃  |     |     |     |     |
| 27 m₅ applies w₄, rejected, w₄ removed by m₄                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 28 m₅ applies w₄, rejected, w₄ pref removed by m₅                      | u₅  | w₄  | w₃  | u₇  |     |     |     |
| 29 m₅ inactive                                                        | u₅  | w₄  | w₃  |     |     |     |     |
| 30 m₃ applies w₅, rejected, w₅ removed by m₃                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 31 m₃ applies w₆, rejected, w₆ removed by m₃                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 32 m₁ applies w₂, rejected, w₂ removed by m₁                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 33 m₁ applies w₄, rejected, w₄ removed by m₁                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 34 m₁ advantaged                                                       | u₅  | w₄  | w₃  |     |     |     |     |
| 35 m₁ applies w₅, rejected, w₅ removed by m₁                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 36 m₁ applies w₆, rejected, w₆ removed by m₁                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 37 m₃ applies w₂, rejected, w₂ removed by m₃                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 38 m₁ applies w₄, rejected, w₄ removed by m₁                           | u₅  | w₄  | w₃  |     |     |     | u₇  |
| 39 m₁ inactive                                                        | u₅  | w₄  | w₃  |     |     |     | u₇  |

**Table 5** Trace of running Király's HRT 3/2-approximation algorithm to the maximum stable matching for instance I'' in Figure 3c.
Student preferences: 
\( s_1: p_3 \)
\( s_2: p_4 p_1 p_2 \)
\( s_3: p_3 \)
\( s_4: (p_2 p_3) p_4 p_1 \)

Project details:
\( p_1: \) lecturer \( l_1, c_1 = 2 \)
\( p_2: \) lecturer \( l_1, c_2 = 2 \)
\( p_3: \) lecturer \( l_2, c_3 = 2 \)
\( p_4: \) lecturer \( l_2, c_4 = 1 \)

Lecturer preferences:
\( l_1: s_2 s_4 \quad d_1 = 2 \)
\( l_2: s_4 (s_1 s_2 s_3) d_2 = 2 \)

Resident preferences:
\( r_1: h_3 \)
\( r_2: h_4 h_1 h_2 \)
\( r_3: h_3 \)
\( r_4: (h_2 h_3) h_4 h_1 \)
\( r_5: (h_1 h_2) \)
\( r_6: (h_1 h_2) \)
\( r_7: (h_3 h_4) \)

Hospital preferences:
\( h_1: (r_5 r_6) r_2 r_4 e_1 = 2 \)
\( h_2: (r_5 r_6) r_2 r_4 e_2 = 2 \)
\( h_3: r_7 r_4 (r_1 r_3) e_3 = 2 \)
\( h_4: r_7 r_4 r_2 \quad e_4 = 1 \)

Women’s preferences:
\( w_1: (m_3 m_7) \)
\( w_2: m_4 (m_1 m_5) (m_2 m_6) \)
\( w_3: (m_3 m_7) \)
\( w_4: (m_2 m_6 m_3 m_7) m_4 (m_1 m_5) \)
\( w_5: (m_1 m_5 m_2 m_6) \)
\( w_6: (m_1 m_5 m_2 m_6) \)
\( w_7: (m_3 m_4 m_7) \)

Men’s preferences:
\( m_1: (w_5 w_6) w_2 w_4 \)
\( m_2: (w_5 w_6) w_2 w_4 \)
\( m_3: w_7 w_4 (w_1 w_3) \)
\( m_4: w_7 w_4 w_2 \)
\( m_5: (w_5 w_6) w_2 w_4 \)
\( m_6: (w_5 w_6) w_2 w_4 \)
\( m_7: w_7 w_4 (w_1 w_3) \)

\( \text{(a) Example SPA-ST instance } I. \)
\( \text{(b) \hbox{\textsf{hrt}} instance } I' \text{ converted from the SPA-ST instance in Figure 3a.} \)
\( \text{(c) \hbox{\textsf{smt}} instance } I'' \text{ converted from the \hbox{\textsf{hrt}} instance in Figure 3b.} \)

Figure 3 Conversion of an \hbox{\textsf{spa-st}} instance to an \hbox{\textsf{smt}} instance.

being able to ever apply to \( w_4 \) as an advantaged man and the benefits of having \( m_2 \) and \( m_3 \) tied at the beginning of \( w_4 \)’s preference list are not realised.

\section*{B \( \frac{3}{2} \)-approximation algorithm correctness proofs}

\subsection*{B.1 Preliminary proofs}

\textbf{Proposition 6.} Let \( T_0 \) denote the point in Algorithm 1’s execution at the end of the main while loop. If a project \( p_j \) is not fully available at some point before \( T_0 \), then it cannot subsequently become fully available before \( T_0 \).

\textbf{Proof.} Assume for contradiction that project \( p_j \) is not fully available at some point before \( T_0 \), but then subsequently becomes fully available before \( T_0 \). At a point where \( p_j \) is not fully available, either \( p_j \) is full or \( l_k \) is full (or both), where \( l_k \) offers \( p_j \). If \( l_k \) is full, it is clear that \( l_k \) must remain so, since students can only be removed from a project of \( l_k \)’s if they are immediately replaced by another student assigning to a project of \( l_k \). Therefore assume that \( p_j \) is full. Then they must somehow become undersubscribed in order to be classified as fully available. The only way this can happen before \( T_0 \) is if lecturer \( l_k \) removes a student assigned to \( p_j \) in order to replace them with a student becoming assigned to another project of \( l_k \)’s. But then this deletion can only occur if \( l_k \) is full and as above \( l_k \) remains full, so \( p_j \) cannot become fully available before \( T_0 \), a contradiction.

\textbf{Proposition 7.} Suppose a blocking pair \((s_i, p_{j'})\) of type (3bi) exists at the end of the main while loop of Algorithm 1, where \( l_k \) offers \( p_{j'} \), and denote this time by \( T_0 \). Then at time \( T_0 \), \( l_k \) is full.
Students preferences: $s_1: p_3$
$s_2: p_4, p_1, p_2$
$s_3: p_3$
$s_4: (p_2, p_3, p_4, p_1)$

Resident preferences: $r_1: h_3$
$r_2: h_4, h_1, h_2$
$r_3: h_3$
$r_4: (h_2, h_3, h_4, h_1)$
$r_5: (h_3, h_4)$

Women’s preferences:

$w_1: (m_3, m_5)$
$w_2: m_4, m_1, m_2$
$w_3: (m_3, m_5)$
$w_4: (m_2, m_3, m_5, m_4, m_1)$
$w_5: (m_3, m_5, m_4)$

Project details:

$p_1$: lecturer $l_1$, $c_1 = 1$
$p_2$: lecturer $l_1$, $c_2 = 1$
$p_3$: lecturer $l_2$, $c_3 = 2$
$p_4$: lecturer $l_2$, $c_4 = 1$

Lecturer preferences:

$l_1$: $s_2, s_4$
$d_1 = 2$
$l_2$: $s_4, (s_1, s_2, s_3)$
$d_2 = 2$

Hospital preferences:

$h_1: r_2, r_4$
$e_1 = 1$
$h_2: r_2, r_4$
$e_2 = 1$
$h_3: r_5, r_4, (r_1, r_3)$
$e_3 = 2$
$h_4: r_5, r_4, r_2$
$e_4 = 1$

Men’s preferences:

$m_1: w_2, w_4$
$m_2: w_2, w_4$
$m_3: w_5, w_4, (w_1, w_3)$
$m_4: w_5, w_4, w_2$
$m_5: w_5, w_4, (w_1, w_3)$

(a) Example SPA-ST instance
(b) HRT instance $I''_2$ converted from the SPA-ST instance in Figure 4a
(c) SMTI instance $I''_2$ converted from the HRT instance in Figure 4b

Figure 4 Conversion of an SPA-ST instance to an SMTI instance.

| Action | $m_1$ | $m_2$ | $m_3$ | $m_4$ | $m_5$ |
|--------|-------|-------|-------|-------|-------|
| 1      | $m_5$ w_5, accepted |       |       |       | $w_5$ |
| 2      | $m_4$ w_5, rejected, $w_5$ pref removed by $m_4$ |       |       |       | $w_5$ |
| 3      | $m_3$ w_4, accepted |       |       |       | $w_4$ |
| 4      | $m_3$ w_5, rejected, $w_5$ pref removed by $m_3$ |       |       |       | $w_4$ |
| 5      | $m_3$ w_4, accepted, $w_4$ pref removed by $m_4$ |       |       | $w_4$ | $w_5$ |
| 6      | $m_4$ w_2, accepted |       |       |       | $w_4$ |
| 7      | $m_2$ w_2, rejected, $w_2$ pref removed by $m_2$ |       |       |       | $w_4$ |
| 8      | $m_2$ w_4, rejected, $w_4$ pref removed by $m_2$ |       |       |       | $w_4$ |
| 9      | $m_2$ advantaged |       |       |       | $w_4$ |
| 10     | $m_2$ w_2, rejected, $w_2$ pref removed by $m_2$ |       |       |       | $w_4$ |
| 11     | $m_2$ w_4, accepted, $w_4$ pref removed by $m_3$ |       |       |       | $w_4$ |
| 12     | $m_3$ w_1, accepted |       |       |       | $w_4$ |
| 13     | $m_1$ w_2, rejected, $w_2$ pref removed by $m_1$ |       |       |       | $w_4$ |
| 14     | $m_1$ w_4, rejected, $w_4$ pref removed by $m_1$ |       |       |       | $w_4$ |
| 15     | $m_1$ advantaged |       |       |       | $w_4$ |
| 16     | $m_1$ w_2, rejected, $w_2$ pref removed by $m_1$ |       |       |       | $w_4$ |
| 17     | $m_1$ w_4, rejected, $w_4$ pref removed by $m_1$ |       |       |       | $w_4$ |
| 18     | $m_1$ inactive |       |       |       | $w_4$ |

Table 6 Trace of running Király’s SMTI 3/2-approximation algorithm to the maximum stable matching for instance $I''_2$ in Figure 5c.
Proof. Assume for contradiction that $l_k$ is undersubscribed at $T_0$. We know that once a
lecturer is full they must remain full (since we can only remove a pair associated with a
lecturer if we are immediately replacing it with an associated pair). Therefore $l_k$ must
have always been undersubscribed. At $T_0$, $p_j$ must be undersubscribed for $(s_i,p_{j'})$ to be
a blocking pair of type (3bi). Therefore at $T_0$, $p_{j'}$ is fully available and must always have
previously been fully available by Proposition 6. But $s_i$ must have applied to $p_{j'}$ at least once
before $T_0$ and as $p_{j'}$ was fully available this must have been accepted. Then since $(s_i,p_{j'})$
is not in the matching at $T_0$ it must have been removed before $T_0$. But in order for this
to happen either $p_{j'}$ or $l_k$ would have to be full, contradicting the fact that $p_{j'}$ was fully
available before $T_0$. Hence $l_k$ must be full at $T_0$. ◀

The following proposition extends Proposition 6.

Proposition 8. During the execution of Algorithm 3, if a project $p_j$ is not fully available
at some point, then it cannot subsequently become fully available.

Proof. Let $T_0$ denote the point in the algorithm’s execution at the end of the main while
loop. We know from Proposition 6 that if a project is not fully available before $T_0$ then it
cannot subsequently become fully available before $T_0$.

Let lecturer $l_k$ offer project $p_j$ and assume project $p_j$ is not fully available at $T_0$. If $l_k$
contains no blocking pairs of type (3bi) then there can be no changes to allocations of $p_j$
after $T_0$. Therefore assume $l_k$ contains at least one blocking pair. Then by Proposition 4
$l_k$ is full at $T_0$. But Algorithm 3 does not change the student allocations for any lecturer and
hence $l_k$ remains full and $p_j$ cannot subsequently become fully available.

It remains to show that if $p_j$ is fully available at $T_0$, that it cannot subsequently cease to be
fully available and then return to be fully available before the end of the algorithms execution.
Since $l_k$ is undersubscribed at $T_0$, $l_k$ cannot contain any blocking pairs by Proposition 4.
Since Algorithm 3 can only affect allocations of projects offered by a full lecturer it is not
possible for $p_j$ to change to being not fully available after $T_0$. ◀

Proposition 9. No student promotion carried out in Algorithm 3 can be to a fully available
project or create a precarious pair.

Proof. Suppose in Algorithm 3 $s_i$ is being promoted from project $p_j$ to project $p_{j'}$ both
offered by lecturer $l_k$. We know that in the main while loop $s_i$ must have iterated over their
preference list at least to the position of $p_j$ (and perhaps further if $s_i$ has been previously
promoted). Therefore, $s_i$ has either been removed from and / or rejected by all projects at
the same rank as $p_{j'}$ in their preference list at least once. This can only occur if each of
those projects was not fully available at the time and by Proposition 8 none of these projects
could subsequently be fully available. Therefore when $s_i$ is promoted to $p_{j'}$, it can never
form a precarious pair. ◀

Proposition 10. Let lecturer $l_k$ offers project $p_j$. If $l_k$ is full and not precarious at some
point, then they cannot subsequently become precarious. Similarly, if a project $p_j$ is full and
not precarious at some point, then it cannot subsequently become precarious. Further if $l_k$
is full and $p_j$ is not precarious then $p_j$ cannot subsequently become precarious.

Proof. We know that a precarious pair cannot be created in Algorithm 3 by Proposition 9.
therefore we focus on the main while loop of Algorithm 4. Let lecturer $l_k$ be full and not
precarious at some point during the main while loop Algorithm 4’s execution and assume
that they later becomes precarious. Since $l_k$ is currently not precarious, the only way they
can become so is by a student $s_i$ forming a precarious assignment to a project of $l_k$’s. But recall that a student will first apply to fully available projects at the head of their preference list. Since $l_k$ is full, no project of $l_k$’s can be considered fully available. In order for $s_i$ to apply to a project of $l_k$’s they must first apply to all fully available projects at the head of their list, gain the assignment and then be removed from $M$. But if a pair $(s_i, p_j)$ is removed from $M$, $p_j$ cannot be fully available. Ultimately, $s_i$ will have exhausted all previously fully available projects at the head of their list before eventually applying to a project of $l_k$’s. But then at that point $s_i$ cannot be precarious giving a contradiction.

Now, let $p_j$ be full and not precarious at some point during the main while loop of Algorithm 1’s execution and assume that it later becomes precarious. If $p_j$ remains full until the time at which it becomes precarious then using similar reasoning to above, any student assigning to $p_j$ cannot be precarious giving a contradiction. If $p_j$ becomes undersubscribed at any point then $l_k$ must be full for the remainder of the algorithm by Proposition 8. It is possible at this point that $l_k$ is precarious (with precarious pairs that include projects other than $p_j$) but since $l_k$ is full, $p_j$ is not fully available. Therefore using similar reasoning to before, any student assigning to $p_j$ cannot be precarious giving a contradiction.

It also follows then that if $l_k$ is full and $p_j$ is not precarious then $p_j$ cannot subsequently become precarious.

\begin{proposition}
Suppose a blocking pair $(s_i, p_j)$ of type (3b) exists at the end of the main while loop of Algorithm 4, where $l_k$ offers $p_{j'}$, and denote this time by $T_0$. Then at time $T_0$, $l_k$ is non-precarious.
\end{proposition}

\begin{proof}
Let $M_0$ be the matching being built at $T_0$ and let $(s_i, p_j) \in M_0$ with $l_k$ offering $p_j$. Suppose for contradiction that $l_k$ is precarious at $T_0$.

As $(s_i, p_{j'})$ is a blocking pair of type (3b), $p_{j'}$ must be undersubscribed at $T_0$. Also, since $(s_i, p_{j'})$ is a blocking pair we know that $s_i$ prefers $p_{j'}$ to $p_j$. Therefore $s_i$ must have removed $p_{j'}$ from their preference list. Denote this time as $T_1$. The removal at $T_1$ occurred either because $(s_i, p_{j'})$ was removed as a non-precarious pair, or because $s_i$ was rejected on application to $p_{j'}$.

- If $(s_i, p_{j'})$ was removed as a non-precarious pair at $T_1$ then either $l_k$ was full, non-precarious and $s_i$ was a worst student assigned to $l_k$, or $p_{j'}$ was full, non-precarious and $s_i$ was a worst student assigned to $p_{j'}$.

- If on the other hand, $s_i$ was rejected on application to $p_{j'}$ at $T_1$, we know that either $l_k$ was full, non-precarious and $l_k$ did not meta-prefer $s_i$ to a worst student in $M(l_k)$, or $p_{j'}$ was full, non-precarious and $l_k$ did not meta-prefer $s_i$ to a worst student in $M(p_j)$.

Whichever of these possibilities occurred we know that at $T_1$, either $l_k$ was full and non-precarious or $p_{j'}$ was full and non-precarious.

- Firstly suppose $l_k$ was full and non-precarious at $T_1$. In this case by Proposition 10, $l_k$ cannot subsequently become precarious, a contradiction to the fact that $l_k$ is precarious at $T_0$.

- Therefore, $p_{j'}$ must have full and non-precarious at $T_1$. By Proposition 10, $p_{j'}$ cannot subsequently become precarious. We also know that $p_{j'}$ must go from being full to being undersubscribed since $p_{j'}$ is undersubscribed at $T_0$. Denote this point in the algorithm’s execution as $T_2$. At $T_2$, $p_{j'}$ must be non-precarious by above and so a non-precarious pair involved with $p_{j'}$ is removed and replaced with a pair involved with some other project of $l_k$’s. This could only happen if $l_k$ was full and non-precarious and so as before cannot again become precarious, a contradiction.
Therefore, $l_k$ must be non-precarious at $T_0$.

**Proposition 12.** Algorithm 3 cannot change the fully available or precarious status of any project or lecturer. Additionally, Algorithm 3 cannot change the precarious status of any pair.

**Proof.** By Proposition 9 in Algorithm 3 it is not possible to assign a student to a fully available project. Therefore we cannot change a fully available project or lecturer to being not fully available. Also, any promotions that take place will be to remove a blocking pair of type (3b), and so by definition the lecturer involved, say $l_k$, will be full and $|M(l_k)|$ will remain the same. Therefore, none of $l_k$’s projects are not fully available at the start of Algorithm 3 and must remain so.

By Proposition 9 in Algorithm 3 it is not possible to create a precarious pair, meaning we cannot change a non-precarious project or lecturer to being a precarious project or lecturer. Finally, by Proposition 11 no changes can be made to any assignments involving a precarious project or lecturer, hence we cannot change a precarious project or lecturer to be non-precarious. Also, since Algorithm 3 cannot change the fully available status of any project, it is not possible for a pair to change it’s precarious status.

Recall, a worse student than student $s_i$, according to lecturer $l_k$, is any student with a lower rank than $s_i$ on $l_k$’s preference list, or if $s_i$ is in phase 2, any student of the same rank that is in phase 1.

**Proposition 13.** Let $T_0$ denote the point in Algorithm 1’s execution at the end of the main while loop. Then the following statements are true.

1. If a project $p_j$ offered by $l_k$ is full before $T_0$, then a student $s_i$ worse than or equal to $l_k$’s worst ranked assignee(s) in $M(p_j)$ cannot subsequently become assigned to $p_j$ before $T_0$ unless $p_j$ is precarious when $s_i$ applies to $p_j$.

2. If a lecturer $l_k$ is full, then a student $s_i$ worse than or equal to $l_k$’s worst assignee(s) cannot become assigned to a project $p_j$ offered by $l_k$ unless this occurs during the main while loop and $l_k$ is precarious when $s_i$ applies to $p_j$.

**Proof.** We deal with each case separately.

1. Assume first that $p_j$ is full and not precarious. Before $T_0$, student $s_i$ can only be added to the matching if the conditions on Line 17 are held, and $l_k$ meta-prefers $s_i$ to a worst assignee in $M(p_j)$. Therefore $l_k$ cannot accept a student that is worse than or equal to a worst student in $M(p_j)$ before $T_0$.

2. It is clear that in Algorithm 3 students assigned to a particular lecturer cannot change, hence we concentrate only on the main while loop of Algorithm 1. Now assume that $l_k$ is full and not precarious before $T_0$. Also since the case where $p_j$ is full has been dealt with, assume that $p_j$ is undersubscribed. Since $l_k$ is not precarious, if $s_i$ were to be added to the matching, $l_k$ must adhere to the conditions on Line 3 and so meta-prefer $s_i$ to a worst assignee of $l_k$. Therefore $l_k$ cannot accept a student that is worse than or equal to a worst student in $M(l_k)$.

Therefore each case is proved.

If on the other hand, as an example, $l_k$ is full and precarious (or $p_j$ is full and precarious) before $T_0$ it is easy to see that a student worse than the current lowest ranked assignee in $M(l_k) \setminus M(p_j)$ could be added before $T_0$. 

B.2 Stability

Lemma 14. Let $M_1$ denote the matching constructed immediately after the main while loop in Algorithm 1 has completed and let $T_1$ denote this point in the algorithm’s execution. At $T_1$, no blocking pair of type (3a), (3bii) or (3c) can exist relative to $M_1$.

Proof. Assume for contradiction that $(s_{b_1}, p_{b_2})$ is a blocking pair of $M_1$ of type (3a), (3bii) or (3c). Let $b_3$ be the lecturer who offers $p_{b_2}$.

It must be the case that in $M_1$, $s_{b_1}$ is either assigned to a project of lower rank than $p_{b_2}$ or is assigned to no project. Therefore, $s_{b_1}$ must have removed $p_{b_2}$ from their preference list during the main while loop of Algorithm 1. Let $M_0$ denote the matching constructed immediately before $p_{b_2}$ was first removed from $s_{b_1}$’s list and let $T_0$ denote this point in the algorithm’s execution. We know that $p_{b_2}$ cannot be fully available at $T_0$ (otherwise $(s_{b_1}, p_{b_2})$ would have been added to $M_0$) and cannot subsequently become fully available by Proposition 8. There are three places where $p_{b_2}$ could be removed from $s_{b_1}$’s list, namely Lines 13, 22 and 27. We look at each type of blocking pair in turn.

- (3a) - Assume we have a blocking pair of type (3a) in $M_1$. Then, $p_{b_2}$ and $l_{b_3}$ are both undersubscribed (and hence $p_{b_2}$ is fully available) in $M_1$. But this contradicts the above statement that $p_{b_2}$ cannot be fully available after $T_0$.

- (3bii) & (3c) - Assume we have a blocking pair of type (3bii) or (3c) in $M_1$. At $T_0$, $p_{b_2}$ is not fully available and so either $p_{b_2}$ is undersubscribed with $l_{b_3}$ being full, or $p_{b_2}$ is full.

- If $p_{b_2}$ was undersubscribed and $l_{b_3}$ was full at $T_0$ then $l_{b_3}$ cannot have been precarious (since $s_{b_1}$ is about to remove $p_{b_2}$ from their list) and by Proposition 10 cannot subsequently become precarious. By Proposition 13, $l_{b_3}$ cannot subsequently accept a student ranked lower than a worst student in $M_0(l_{b_3})$. Also either $s_{b_1}$ is a worst assignee in $M_0(l_{b_3})$ (Line 13), or $l_{b_3}$ ranks $s_{b_1}$ at least as badly as a worst student in $M_0(l_{b_3})$ (Line 27). Lecturer $l_{b_3}$ must remain full for the rest of the algorithm since if a student is removed from a project offered by $l_{b_3}$ then they are immediately replaced. Therefore $l_{b_3}$ must be full in $M_1$, and since $l_{b_3}$ cannot have accepted a worse ranked student than $s_{b_1}$ after $T_0$, $(s_{b_1}, p_{b_2})$ cannot be a blocking pair of $M_1$ of type (3bii) or (3c).

- Instead assume at $T_0$ that $p_{b_2}$ is full in $M_0$. As $s_{b_1}$ is about to remove $p_{b_2}$, we know $p_{b_2}$ cannot be precarious, and by Proposition 10 cannot subsequently become precarious. Either $s_{b_1}$ is a worst assignee in $M_0(p_{b_2})$ (Line 22), or $l_{b_3}$ ranks $s_{b_1}$ at least as badly as a worst student in $M_0(p_{b_2})$ (Line 27).

If $p_{b_2}$ remains full until $T_1$, then clearly $(s_{b_1}, p_{b_2})$ cannot block $M_1$ by Proposition 13. So assume $p_{b_2}$ becomes undersubscribed at some point between $T_0$ and $T_1$ for the first time, say at $T_{0.5}$. This can only happen if $l_{b_3}$ is full and there is a student $s_i$ who assigns to another project $p_j$ that $l_{b_3}$ offers, where $s_i$ is meta-preferred to a worst student in $M_0(l_{b_3})$. This worst student must also be a worst student in $M_0(p_{b_2})$ since we are removing from $M_0$ a pair associated with $p_{b_2}$. But then $s_{b_1}$ must be ranked at least as badly as a worst student in $M_0(l_{b_3})$. Using similar reasoning to the previous case, $l_{b_3}$ must be full in $M_1$, non-precarious, and since $l_{b_3}$ cannot have accepted a worse ranked student than $s_{b_1}$ after $T_{0.5}$, $(s_{b_1}, p_{b_2})$ cannot be a blocking pair of $M_1$ of type (3bii) or (3c).

Hence it is not possible for $(s_{b_1}, p_{b_2})$ to be a blocking pair of $M$ of type (3a), (3bii) or (3c) after the main while loop.
Proposition 15. Let $M_1$ be the matching constructed immediately at the end of the main while loop of Algorithm 1’s execution, and let $T_1$ denote this point in the algorithm’s execution. At $T_1$, for each blocking pair $(s_i, p_j)$ of type (3bi), $s_i$ must be one of the worst assignees of $M_1(l_k)$, where $l_k$ offers $p_j$.

Proof. Since $(s_i, p_j)$ is a blocking pair of type (3bi), we know that $s_i$ is assigned to another project, say $p_j'$, of $l_k$’s in $M_1$, where $s_i$ prefers $p_j'$ to $p_j$.

During the main while loop’s execution, student $s_i$ must have removed $p_j'$ from their preference list in order to eventually assign to $p_j$. Student $s_i$’s removal of $p_j'$ could only happen if $p_j'$ was not precarious at this point. By Proposition 10, $p_j'$ cannot subsequently become precarious. Let the matching constructed immediately before this removal be denoted by $M_0$ and let $T_0$ denote this point in the algorithm’s execution. Project $p_j'$ was either full or undersubscribed at $T_0$. We show that in both of these cases, $s_i$ is one of the worst students in $M_1(l_k)$.

Suppose $p_j'$ was full at $T_0$. Then as $p_j'$ is not precarious, $p_j'$ cannot subsequently be assigned a student worse than the worst assignee in $M_0(p_j')$ up until $T_1$, by Proposition 13. Since $s_i$ removed $p_j'$ from their list while $p_j'$ was full we know that either $s_i$ is a worst assignee in $M_0(p_j')$ (Line 22), or $l_k$ ranks $s_i$ at least as badly as a worst student in $M_0(p_j')$ (Line 27). Between $T_0$ and $T_1$, $s_i$ assigns to the project $p_j$, at a worse rank than $p_j'$ in $s_i$’s list, where $p_j$ is also offered by $l_k$.

Now, we know that $p_j'$ becomes undersubscribed by $T_1$ and so it must be the case that there is a point $T_{0.5}$ between $T_0$ and $T_1$, such that another student $s_i$ assigns to a project (not $p_j'$) of $l_k$’s which removes pair $(s_{i''}, p_{j''})$ from the matching constructed just before that removal, denoted by $M_{0.5}$. Let $T_{0.5}$ be the first point at which $p_j'$ becomes undersubscribed after $T_0$. Lecturer $l_k$ must be full at this point since the addition of $s_{i''}$ removes a student (namely $s_{i''}$) from a different project (namely $p_{j''}$). Also, lecturer $l_k$ cannot have been precarious, otherwise $p_j'$ would have been identified as a precarious project at Line 10, but we know $p_j'$ cannot have been precarious after $T_0$. So $s_{i''}$ must have been a worst assignee in $M_{0.5}(l_k)$ and therefore $M_{0.5}(p_{j''})$. Since $T_{0.5}$ is the first point $p_{j''}$ becomes undersubscribed after $T_0$, the worst student in $M_{0.5}(p_{j''})$ can be no worse than the worst student in $M_{0.5}(p_{j'})$ by Proposition 13. Student $s_i$ must be either a worst student in $M_{0.5}(l_k)$, or be as bad as a worst student in $M_{0.5}(l_k)$. By Propositions 10 and 13, $l_k$ cannot be assigned a worse student than $s_i$ between $T_{0.5}$ to $T_1$, and so as $s_i$ is assigned to $l_k$ at $T_1$ then they must be one of the worst students in $M_1(l_k)$.

Suppose then that $p_j'$ is undersubscribed at $T_0$. Since a preference element is being removed, $l_k$ must have been full, non-precarious and either $s_i$ is a worst assignee in $M_0(l_k)$ (Line 13), or $l_k$ ranks $s_i$ at least as badly as a worst student in $M_0(l_k)$ (Line 27). But then by Propositions 10 and 13, $l_k$ must have remained non-precarious until $T_1$ and been unable to assign to a worse student than $s_i$. Since $s_i$ is assigned to a project of $l_k$’s, $s_i$ must be one of the worst students in $M_1(l_k)$.

Therefore, for any blocking pair $(s_i, p_{j'})$ of type (3bi) of $M_1$, $s_i$ must be one of the worst students in $M_1(l_k)$.

Proposition 16. In Algorithm 3, if a blocking pair $(s_{i''}, p_j)$ of type (3bi) is created (in the process of removing a different blocking pair of type (3bi)) then $s_{i''}$ must be one of the worst students in $M_2(l_k)$, where $l_k$ is the lecturer who offers $p_j$ and $M_2$ is the matching constructed immediately after this removal occurs.
Proof. Let $M_0$ denote the matching at the end of the main while loop of Algorithm 1 and let $T_0$ denote this point in the algorithm’s execution. Assume that during Algorithm 3’s execution, the first promotion to reveal a blocking pair $(s_i, p_j)$ of type $3(bii)$ occurs such that $s_i$ is not a worst student in $M_1(l_k)$. Let $M_1$ be the matching constructed just before this promotion occurs. Suppose the promotion involves student $s_i$ moving from a less preferred $p_j$ to a more preferred project $p_{j'}$. It is clear that in the removal of blocking pairs of type $(3bi)$ there is no change in regards to which students are assigned to projects of $l_k$, therefore the same students are assigned to each lecturer in $M_0$, $M_1$ and $M_2$. By Proposition 15, $s_i$ is and remains one of the worst assignees of $l_k$ in $M_0$, $M_1$ and $M_2$. Since $s_{i'}$ is not a worst assignee in $M_2(l_k)$, $l_k$ must prefer $s_{i'}$ to $s_i$. But this would mean $(s_{i'}, p_j)$ was a blocking pair of type $(3c)$ in $M_0$ if $p_j$ were full or $(3bii)$ in $M_0$ if $p_j$ were undersubscribed, both a contradiction to Lemma 14.

$\blacktriangleright$ Proposition 17. It is not possible in Algorithm 3’s execution, for a blocking pair of any type other than $(3bi)$ to be created.

Proof. Let $M_0$ denote the matching constructed immediately after the main while loop of Algorithm 1 terminates and let $T_0$ denote this point in the algorithm’s execution. By Lemma 14 only blocking pairs of type $(3bi)$ of $M_0$ may exist at $T_0$ therefore we restrict our attention to the removal of such pairs. Assume for a contradiction that during Algorithm 3’s execution, the first promotion to reveal a blocking pair of type not equal to $(3bi)$ occurs. Let $M_1$ (respectively $M_2$) be the matching constructed just before (respectively after) this promotion occurs with $T_1$ (respectively $T_2$) denoting this point in the algorithm’s execution. Suppose that this promotion involves student $s_i$ being promoted from project $p_j$ to project $p_{j'}$ as pair $(s_i, p_{j'})$ is a blocking pair of $M_1$ of type $(3bi)$. Since $(s_i, p_{j'})$ is a blocking pair of $M_1$ of type $(3bi)$ we know that $p_j$ and $p_{j'}$ are both offered by the same lecturer, say $l_k$. Assume that this promotion has now revealed a blocking pair $(s_{i'}, p_j)$ of type $(3a)$, $(3bii)$ or $(3c)$ in $M_2$. We look at each case in turn.

- $(3a)$ - Since in $M_1$, $(s_i, p_{j'})$ was a blocking pair of type $(3bi)$ we know that $l_k$ is full at $T_1$. The promotion involves moving $s_i$ from one project offered by $l_k$ to another, therefore at $T_2$, $l_k$ must be full and so $p_j$ cannot be involved in a blocking pair of type $(3a)$ in $M_2$, a contradiction.

- $(3bii)$ - Suppose $(s_{i'}, p_j)$ is a blocking pair of type $(3bii)$ in $M_2$. Since it is of type $(3bii)$, $l_k$ must prefer $s_{i'}$ to a worst assignee in $M_2(l_k)$ (and consequently $M_1(l_k)$ as students do not change lecturer). If $p_j$ was undersubscribed at $T_1$ then $(s_{i'}, p_j)$ would have constituted a blocking pair of type $(3bii)$, a contradiction. Therefore $p_j$ must have been full in $M_1$. We know that $(s_{i'}, p_j) \in M_1$ and that $s_i$ is a worst assignee in $M_1(l_k)$ by Proposition 15 and 16 therefore $l_k$ prefers $s_{i'}$ to $s_i$. It follows that $(s_{i'}, p_j)$ would have constituted a blocking pair in $M_1$ of type $(3c)$, a contradiction to the fact that no blocking pair of any type other than $(3bi)$ was revealed prior to $T_2$.

- $(3c)$ - Suppose finally that $(s_{i'}, p_j)$ is a blocking pair of $M_2$ of type $(3c)$. But blocking pairs of type $(3c)$ require $p_j$ to be full in $M_2$ which it cannot be since $(s_i, p_j)$ has been removed just before $T_2$, hence $p_j$ cannot be involved in a blocking pair of type $(3c)$. Therefore it is not possible for a blocking pair of type $(3a)$, $(3bii)$ or $(3c)$ to be created during the first promotion of a student, and hence any promotion.

$\blacktriangleright$ Theorem 18. Any matching produced by Algorithm 1 must be stable.

Proof. Let $M_0$ be the matching constructed immediately after the termination of the main while loop of Algorithm 1 and let $T_0$ denote this stage of the algorithm. Recall that by
Lemma 14 only blocking pairs of type \((3b)\) may exist relative to \(M_0\). Also, by Lemma 17 no blocking pair of any other type can exist relative to the matching constructed after \(T_0\).

Algorithm 3 systematically removes blocking pairs of type \((3b)\) in a series of student promotions. Each promotion improves the outcome for a student.

Therefore there are no blocking pairs of any type in the finalised matching \(M_s\) and so \(M_s\) is stable.

Since this proof relies only on the fact that \(p_{b2}\) is removed from \(s_b\)’s list once for \((s_b, p_{b2})\) not to become a blocking pair, we can infer that if we allowed students to only iterate once through their preference preference list rather than twice, this would still result in a stable matching.

B.3 Time complexity and termination

Proposition 19. The maximum number of times a student \(s_i\) can apply to a project \(p_j\) on their preference list during the main while loop of Algorithm 7 is three.

Proof. First we note that as soon as \(s_i\) removes \(p_j\) from their preference list once during Algorithm 7’s execution, \((s_i, p_j)\) cannot currently be or subsequently become a precarious pair by Proposition 10.

Focussing on the main while loop, assume for some iteration, that phase 1 student \(s_i\) applies to project \(p_j\) on their preference list. Either \((s_i, p_j)\) is added to the matching being built \(M\), or \(p_j\) is removed from \(s_i\)’s list. If \(p_j\) is removed from \(s_i\)’s list then \(s_i\) may still apply to project \(p_j\) in phase 2 but as noted above \((s_i, p_j)\) cannot become a precarious pair.

Assume instead that \((s_i, p_j)\) is added to \(M\). If it remains in \(M\) until the algorithm completes then \(s_i\) cannot apply to \(p_j\) again. So assume that \((s_i, p_j)\) is removed from \(M\) at some point due to another pair being added to \(M\). If \((s_i, p_j)\) was not precarious at the point it is removed from \(M\) then \(s_i\) removes \(p_j\) from their list and the next time \(s_i\) could apply to \(p_j\) is when \(s_i\) is in phase 2 when as above \((s_i, p_j)\) cannot become a precarious pair.

Assume therefore that \((s_i, p_j)\) was precarious when removed from \(M\). Then \(s_i\) does not remove \(p_j\) from their list and \(s_i\) can again apply to \(p_j\) during phase 1. Note that if \((s_i, p_j)\) is re-added to \(M\) it must be as a non-precarious pair. This is because, using similar reasoning that was used in Proposition 10 at the point at which \(s_i\) reapply to \(p_j\) they must have exhausted all fully available projects at the head of their list, therefore \((s_i, p_j)\) cannot again become precarious. Therefore, \(s_i\) can apply to \(p_j\) a maximum of three times during the execution of the while loop: at most twice while \(s_i\) is in phase 1 (twice only if \((s_i, p_j)\) is removed as a precarious pair) and at most once in phase 2.

Lemma 20. All operations inside the main while loop of Algorithm 7 run in constant time.

Proof. The data structures required are described below and are summarised in Figure 5. For initialisation purposes, each student, project and lecturer has a list of length \(n_2\), \(n_1\) and \(n_3\) respectively, each entry of which requires \(O(1)\) space. In order to not exceed a time complexity of order the sum of lengths of preference lists, a process of virtual initialisation is used on these data structures [8, p. 149].

Student data structures. For each student a doubly-linked list of (project, rank) tuples embedded in an array, prefList, stores their preference list in order of rank, representing the undeleted entries. A small example is shown in Figure 5 with \(p_2\), \(p_2\) and \(p_1\) all of rank 1 on \(s_i\)’s preference list. Entries may be deleted from this array; a copy of this list prior to any deletions being carried out is retained in order to allow a second iteration through a
student’s preference list, if they move into phase 2. An array projPosition of length \( n_2 \) retains links to the position of (project, rank) tuples in prefList, allowing a constant time removal of projects from prefList. An integer variable associated with each student stores which phase this student is in. Examples for these final two data structures are also shown in Figure 5.

**Project data structures.** Each project has a link to their supervising lecturer. An array, projectedPrefList stores the projected preference list of \( l_k \) for \( p_j \) in the form of (student, rank, boolean) tuples. As an example, suppose \( p_j \) has a projected preference list starting with \( s_7 \) at rank 1, \( s_4 \) at rank 1 and \( s_6 \) at rank 2 as is shown in Figure 5. The boolean values indicate which student-project pairs are currently in the matching.

Once a project is full and non-precarious it cannot accept a worse student than it already has for the remainder of the algorithm, according to Proposition 10. Assume \( p_j \) is full and non-precarious. Let the worst student assigned to \( p_j \) be given by \( s_w \). We retain a pointer, last, which points to the rightmost student at the same rank as \( s_w \) in \( p_j \)’s projectedPrefList. This pointer must move from right to left in a linear fashion (moving up in ranks) given the above proposition.

During the course of the algorithm, we may need to remove the worst student according to \( l_k \) from \( M(p_j) \). It is possible that there are two or more students who are worst assignees with some being in phase 1 and some in phase 2. In order to ensure that we prioritise the removal of phase 1 students, two pointers are added for each entry in projectedPrefList, which point to the head of a phase 1 and a phase 2 doubly-linked list associated with that tie embedded in the projectedPrefList array (this data structure is not shown in Figure 5). Adding or removing a phase 1 or 2 student to either list takes constant time, as they do not need to be kept in order. Then, un-assigning a student requires a check to be made in the tie associated with the worst position (found using last), in order to prioritise a phase 1 student’s removal. In total this takes constant time. Note that a student can only change phase if they are not allocated and therefore updating an allocated student’s phase in these lists is not necessary unless they have just been added.

Each project also contains a doubly-linked list embedded in an array of students, denoted by precariousList, containing students who have formed a precarious pair with this project. In the example in Figure 5, \( p_j \) is not precarious and so no students form a precarious pair with \( p_j \). Adding to and removing from this list takes constant time if we assume that \( s_i \) is stored at index \( i - 1 \). A project \( p_j \) supports a student \( s_i \) in being precarious if \((s_i, p_j) \in M \) and \( p_j \) is the first fully available project at the same rank as \( p_j \)’s list. Then, a doubly-linked list embedded in an array of students, supportList, stores the students for which \( p_j \) gives their support. As before, adding to and removing from this list takes constant time. A counter stores the number of students assigned to \( p_j \) in \( M \).

**Lecturer data structures.** For each lecturer an array of (student, rank, boolean) tuples, prefList, stores their preference list in order of rank, with a True value stored in the ith boolean if student \( s_i \) is assigned to a project of \( l_k \)’s. Figure 5 shows an example with \( s_8 \) not assigned to \( l_k \) at rank 1 and \( s_2 \) and \( s_3 \) both assigned at rank 2. Each lecturer also has an array of length \( n_1 \), studentPositions which retains links to the position of (student, rank, boolean) tuples in \( l_k \)’s prefList, and a counter stores the number of students assigned to \( l_k \) in \( M \). A doubly-linked list embedded in an array of projects, precariousProjList, stores the projects offered by \( l_k \) that are precarious, where project \( p_j \) is stored at index \( j - 1 \) in this list if it is precarious. Figure 5 shows \( p_2 \) being a precarious project of \( l_k \).

Similar to projects, by Proposition 10 once a lecturer is full and non-precarious they cannot accept a worse student than they already have for the remainder of the algorithm.
Assume \( l_k \) is full and non-precarious. Using similar data structures described in the Project Data Structures section above we are able to find \( l_k \)'s worst assigned student in constant time.

**Student pointers.** A student \( s_i \) retains two pointers for the project(s) tied at the head of their list. One pointer \textbf{first} stores the first fully available project when iterating from left to right, and \textbf{second} stores the second fully available project. If \textbf{first} (respectively \textbf{second}) reaches the end of the tie, then a boolean \textbf{firstFin} (respectively \textbf{secondFin}) is set to True. For each iteration of the main while loop of Algorithm 1 each student \( s_i \) first seeks a project at the head of their list that is fully available. This will be precisely the project \( p_j \) that \textbf{first} points to. Then if \textbf{second} has not reached the end of the tie, \((s_i, p_j)\) is precarious and the project that \textbf{second} points to, \( p_j' \), supports \( p_j \). If however, \textbf{firstFin} is set to True, then the leftmost project at the head of \( s_i \)'s list is a favourite project, with \((s_i, p_j)\) being unable to become precarious. Proposition 19 shows that maximum number of applications a student can make to a project on their preference list is 3. At most twice in phase 1 (twice if removed as a precarious pair) and once in phase 2.

During phase 2 there are no fully available projects on \( s_i \)'s list since \( s_i \) must have applied and been rejected (in some way) from every project on their preference list at least once already. Therefore, the \textbf{first} and \textbf{second} pointers are not required in phase 2. During phase 1 the the \textbf{first} and \textbf{second} pointers are only required to iterate once over each tie as described above. Hence, the maximum number of times a student’s list is iterated over is 4; once each for the two pointers \textbf{first} and \textbf{second}, once again after \textbf{first} and \textbf{second} have reached the end of the tie at the head of a student’s list (the student may have retained projects at the head of their list after this point if the projects were precarious), and finally once during phase 2.

**Matching data structures.** The current matching is stored in an array of cells \textbf{matchArray} where cell \( i-j \) contains project \( p_j \) if \((s_i, p_j) \in M\) or null otherwise. Figure 5 shows an example with student \( s_2 \) being assigned to project \( p_6 \).

**Processes (in the order encountered in Algorithm 1):**

1. **a student \( s_i \) applies to a favourite project:** if \( s_i \)'s \textbf{FirstFin} is set to True then there are no fully available projects at the head of \( s_i \)'s list, and a favourite project of \( s_i \) will be the leftmost project. If however, \textbf{FirstFin} is False then there are fully available projects at the head of their list and a favourite project of \( s_i \) is pointed to by \textbf{first}, which is retrievable in constant time.
2. **deciding if a project \( p_j \) is undersubscribed or full or deciding if a lecturer \( l_k \) is undersubscribed or full:** Using the counters described above a comparison can be made between \( p_j \)'s capacity and their current number of allocations. A similar comparison can be made for \( l_k \). Both can be achieved in constant time.
3. **deciding if project \( p_j \) is fully available:** \( p_j \) would not be fully available if either \( p_j \) is full or \( l_k \) is full. Therefore a comparison of the number of allocations for \( p_j \) and \( l_k \) and their respective capacities is required. Again this can be achieved in constant time.
4. **adding a pair \((s_i, p_j)\) to \( M\):** Project \( p_j \) is placed in the \( i-j \)th cell of \textbf{matchArray} and \( p_j \) and \( l_k \)'s allocation counters are incremented. Project \( p_j \)'s \textbf{projectedPrefList} and lecturer \( l_k \)'s \textbf{prefList} booleans are updated in constant time using their associated \textbf{studentPositions} data structures. Each tuple in these lists has a link to the head of phase 1 and phase 2 lists for their tie. When the pair is added the tuple is added to either the phase 1 or phase 2 list. If \((s_i, p_j)\) is precarious then \( s_i \) is added to \( p_j \)'s \textbf{precariousList} and the project pointed to by \textbf{second}, \( p_j' \), adds \( s_i \) to their \textbf{supportList}.

If \( p_j \) has just changed from being non-precarious to precarious then \( l_k \) adds \( p_j \) to their
precariousProjList. If the addition of \((s_i, p_j)\) to \(M\) means that \(p_j\) goes from being fully available to not being fully available then we need to ensure that other students who rely on \(p_j\) as their support are updated. Therefore \(p_j\) alerts each student on their supportList that they are no longer to be relied upon as a fully available project. This triggers each of those students to update their second pointers. The time required for this can be attributed to the movement of second pointers as noted earlier. After being alerted, some other pair in \(M\) may stop being precarious, but any changes can be conducted in constant time as described above. If on adding pair \((s_i, p_j)\), \(p_j\) has now become full and non-precarious then the last pointer will move from right to left over projectedPrefList until it reaches the end of a tie whose phase 1 and phase 2 lists are non-empty. From this point on last is only updated upon removing a pair from \(M\) (Point 5).

5. deciding if \(l_k\) is precarious, and returning a precarious project if one exists: Checking whether precariousProjList is empty for \(l_k\) is a simple process that takes constant time. Retrieving a precarious pair should one exist requires selection of the first student from \(l_k\)’s precariousProjList.

6. Finding a worst assignee of \(l_k\) and deciding if \(l_k\) meta-prefers \(s_i\) to this worst assignee: This operation only needs to be executed if \(l_k\) is full and non-precarious. In that situation \(l_k\)’s last pointer will point to the rightmost position in a tie in prefList such that \(l_k\)’s current worst student \(s_w\) is assigned at the same rank. Then as previously discussed, all that is required is to check the links to phase 1 and phase 2 students for this tie and return a phase 1 student if one exists, or phase 2 student if not. This can be conducted in constant time. Deciding if \(l_k\) meta-prefers \(s_i\) to \(s_w\) can also be done in constant time by comparing rank and phase.

7. removing a preference list entry from \(s_i\)’s list: This process is shown in Algorithm 2 which runs in constant time, since we can find a specific project \(p_j\) in \(s_i\)’s prefList using the projPosition array.

8. removing a pair \((s_i, p_j)\) from \(M\): The \(i-1\)’th cell of matchArray is set to null, and \(p_j\) and \(l_k\)’s allocation counters are decremented. Project \(p_j\)’s projectedPrefList and lecturer \(l_k\)’s prefList booleans are updated in constant time. The tuples associated with \(s_i\) in these lists are removed from their phase 1 or phase 2 list in constant time. If a pair \((s_i, p_j)\) is removed from \(M\), then this is either because \(p_j\) or \(l_k\) is full. By Proposition 5 \(p_j\) cannot subsequently become fully available. Thus, the removal of a pair cannot change \(p_j\)’s fully available status. All that is required then is to check whether \((s_i, p_j)\) was precarious, and update \(p_j\)’s precariousList and \(l_k\)’s precariousProjList. If on removing pair \((s_i, p_j)\), the last pointer now points to a tie with empty phase 1 and phase 2 lists, last needs to be updated and accordingly moves from right to left until it reaches the end of a tie with a non-empty phase 1 or phase 2 list.

9. deciding if \(p_j\) is precarious, and returning a precarious pair if one exists: Similar to Point 5 above but using the precariousList of \(p_j\).

10. Finding a worst assignee of \(p_j\) according to \(l_k\) and deciding if \(l_k\) meta-prefers \(s_i\) to this worst assignee: Similar to Point 6 above, this operation is only required if \(p_j\) is full and not precarious, at which point \(p_j\)’s last pointer will point to the rightmost position in a tie in projectedPrefList such that \(l_k\)’s current worst student assigned to \(p_j\), \(s_w\) is assigned at the same rank. As above retrieving \(s_w\) and comparing its rank and phase with \(s_i\) takes constant time.

Therefore, all operations inside the main while loop of Algorithm 1 run in constant time.
Algorithm 5 gives a more detailed look at the processes involved in Algorithm 3. Proposition 21 shows that Algorithm 5 is has linear time complexity.

**Algorithm 5**

Promote-students($M$) Removes all blocking pairs of type 3bi

**Require:** SPA-ST instance $I$ and matching $M$ which does not contain blocking pairs of type (3a), (3bii) or (3c).

**Ensure:** Return a stable matching $M$.

1. create data structures as described in Proposition 21
2. while $S \neq \emptyset$ do
3. pop $p_j$ from stack $S$
4. remove the first student $s_i$ from list $\rho_j$
5. let $\rho_k = M(s_i)$
6. if $s_i$ prefers $p_k$ to $p_j$ then $\triangleright p_j$ is undersubscribed, $s_i$ is assigned and prefers $p_j$ to $M(s_i)$
7. $M \leftarrow M \setminus \{(s_i, p_k)\}$
8. $M \leftarrow M \cup \{(s_i, p_j)\}$
9. let $\rho_k$ be the list of student and rank tuples associated with project $M(s_i)$ and let boolean $\beta_k$ indicate whether $M(s_i)$ is on stack $S$
10. if $\rho_k \neq \emptyset$ then
11. push $p_k$ onto stack $S$ if it is not already on $S \triangleright$ using $\beta_k$
12. end if
13. end if
14. if $\rho_j \neq \emptyset$ and $p_j$ is undersubscribed then
15. push $p_j$ onto stack $S \triangleright p_j$ cannot currently be on $S$
16. end if
17. end while
18. return $M$

**Proposition 21.** The time complexity of Algorithm 5 is $O(m)$ where $m$ is the total length of all students preference lists.

**Proof.** Abraham et al. 4 describes the process of a sequence of promotions for houses in the HA problem in order to return a trade-in-free matching. A similar process is described here to remove all blocking pairs of type (3bi). We create the following data structures.

- A linked list $\rho_j$ of students $s_i$ for each project $p_j$ such that $s_i$ is assigned in $M$, finds $p_j$ acceptable. We may also start by assuming that $\rho_j$ involves only students who prefer $p_j$ to $M(s_i)$, however the time complexity is unaffected by this.
- A ranking list $r_i$ for each student $s_i$ built as an array such that $r_i=j$ contains the rank of $p_j$ for student $s_i$;
- A stack $S$ of undersubscribed projects $p_j$, such that $p_j$ is non-empty, is created;
- A variable $\beta_j$ for each project $p_j$ which records whether $p_j$ is already in $S$.

These data structures can be initialised in $O(m)$ time where $m$ is the total length of all student preference lists.

Execution of Algorithm 5 proceeds as follows. For each iteration of the while loop a project $p_j$ is taken from stack $S$. Project $p_j$ must be undersubscribed and have non-empty list $\rho_j$. The first tuple from $\rho_j$, $(s_i, r)$ is removed and if $s_i$ would prefer to be assigned to $p_j$ than $M(s_i)$ (found by comparing $r$ and $\alpha_i$) then we remove pair $(s_i, M(s_i))$ from $M$ and add
Student $s_i$

$$\begin{array}{|c|c|c|}
\hline
(p_3, 1) & (p_2, 1) & (p_1, 1) \\
\hline
\end{array}$$

---

prefList, (project, rank) tuples

first

second

$$\begin{array}{|c|c|c|}
\hline
3 & 2 & 1 \\
\hline
\end{array}$$

---

projPosition

phase of $s_i$

$F$ firstFin, True if first pointer reaches end of tie

$F$ secondFin, True if second pointer reaches end of tie

Project $p_j$

$$\begin{array}{|c|c|c|}
\hline
(s_7, 1, T) & (s_4, 1, F) & (s_6, 2, T) \\
\hline
\end{array}$$

---

projectedPrefList, (student, rank, boolean) tuples

last

$$\begin{array}{|c|c|c|}
\hline
\text{null} & 5 & \text{null} \\
\hline
\end{array}$$

---

studentPositions

$$\begin{array}{|c|c|c|}
\hline
\text{null} & \text{null} & \text{null} \\
\hline
\end{array}$$

---

precariousList

$$\begin{array}{|c|c|c|}
\hline
\text{null} & \text{null} & s_3 \\
\hline
\end{array}$$

---

supportList

$l_5$ lecturer offering $p_j$

2 number of allocations in $M$

Lecturer $l_k$

$$\begin{array}{|c|c|c|}
\hline
(s_8, 1, F) & (s_2, 2, T) & (s_2, 2, T) \\
\hline
\end{array}$$

---

prefList, (student, rank, boolean) tuples

last

$$\begin{array}{|c|c|c|}
\hline
\text{null} & 2 & 3 \\
\hline
\end{array}$$

---

studentPositions

$$\begin{array}{|c|c|c|}
\hline
\text{null} & p_2 & \text{null} \\
\hline
\end{array}$$

---

precariousProjList

3 number of allocations in $M$

Matching $M$

matchArray, cell $i - 1$ contains $p_j$ if $(s_i, p_j) \in M$ or null if $s_i$ is unmatched

$$\begin{array}{|c|c|c|}
\hline
\text{null} & p_6 & \text{null} \\
\hline
\end{array}$$

---

Key

- boolean/int/link/tuple

- Array

- Doubly linked list embedded in an array

---

Figure 5 Data structures guide for Lemma 20
(s_i, p_j), updating α. Now, M(s_i) is certainly an undersubscribed project and it is added to S (unless it already exists on S). Whether or not (s_i, p_j) is added to M, p_j may still be undersubscribed. If p_j is non-empty and p_j is undersubscribed, then p_j is added to S.

With each iteration we remove a tuple from some project’s p list. These lists must be finite because preference lists are finite and therefore Algorithm 5 will terminate with empty S. It is clear that Algorithm 5 will take O(m) time where m is the total length of all students preference lists.

**Theorem 22.** Algorithm 2 always terminates and runs in linear time with respect to the total lengths of students’ preference lists.

**Proof.** By Proposition 19 each student can apply to a project on their preference list a maximum of three times during the main while loop of Algorithm 1. Since all operations within this while loop run in constant time by Proposition 20, this part of the algorithm must run in O(3m) = O(m) time, where m is the sum of lengths of all students’ preference lists. Proposition 21 shows that Algorithm 5 also runs in O(m) time, therefore so must Algorithm 1. Finally, since student preference lists are of finite length, Algorithm 1 must terminate.

### B.4 Performance guarantee

#### B.4.1 Preliminary definitions

The **underlying graph** G of an [SPA-ST] instance I consists of sets of student, project and lecturer vertices. Edges exist between a student vertex s_i and a project vertex p_j if s_i finds p_j acceptable. Edges exist between a project vertex p_j and lecturer vertex l_k if l_k offers p_j.

The **mapped graph** G’ of the underlying graph G of the [SPA-ST] instance I is created in the following way. Let all student vertices remain unchanged. For each lecturer vertex l_k we create multiple cloned vertices l_k^r_1, ..., l_k^r_k, where r_k = d_k - |M(l_k) ∩ M_opt(l_k)| and d_k is l_k’s capacity. Let M be the matching found by Algorithm 1 for instance I and let M_opt be a maximum stable matching in I. In G there are edges between students and projects, and projects and lecturers, whereas G’ contains only edges between students and lecturer clones.

An (s_i, p_j) edge in G corresponds to an (s_i, l_k^r) edge in G’, where l_k^r denotes the r^{th} lecturer clone of lecturer l_k. Edges in G’ are given by M’ and M’_{opt}, defined below. M’_{opt} edges are defined as follows. For each lecturer l_k, if M_opt(l_k) = {s_i_1, ..., s_i_t} then add (s_i_r, l_k^r), (1 ≤ r ≤ t) to M’_{opt}, the mapped version of M_opt in G’. M’ edges are then added using Algorithm 6. By using this algorithm we ensure that where possible, pairs of edges in M’ and M’_{opt} involving the same project are assigned to the same lecturer clone in G’.

According to Algorithm 6 we do the following. A copy of M\M_{opt} is created and denoted M_0 which intuitively contains the set of student-project pairs that have not yet been mapped. L_0 is a copy of the set of all lecturer clone vertices, and L’_0 is the empty set. Intuitively, L’_0 will collect up any remaining lecturer clones, after pairs of edges in M’ and M’_{opt} involving the same project are dealt with. For each lecturer clone l_k^r ∈ L_0, if there is an edge (s_i, l_k^r) in M’_{opt}, for some s_i then we let p_j be the project assigned to s_i in M_{opt}. If there is not, then l_k^r is added to L’_0. Assuming (s_i, l_k^r) ∈ M’_{opt}, then we check if there is an edge (s_{i_0}, l_k^r) in M_0 for some student s_{i_0}. Again, if there is not then l_k^r is added to L’_0. If (s_{i_0}, p_j) ∈ M_0 for some student s_{i_0} then we add edge (s_{i_0}, l_k^r) to M’ and remove (s_{i_0}, p_j) from M_0. After all lecturer clones have been tested, then for each student-project pair (s_i, p_j) remaining in M_0 we find an unused lecturer clone l_k^r ∈ L’_0, where l_k offers p_j, and add (s_i, l_k^r) to M’. Project vertices and all other edges are ignored in G’.
Algorithm 6 Create-mapped(M), obtains a set of edges $M'$ for the mapped graph $G'$ corresponding to edges in $M \setminus M_{\text{opt}}$

**Require:** An instance $I$ of $\text{SPA-ST}$, a stable matching $M$ and maximum stable matching $M_{\text{opt}}$ of $I$ and a mapped version $M'_{\text{opt}}$ of $M_{\text{opt}}$.

**Ensure:** Return a mapped version $M'$ of $M \setminus M_{\text{opt}}$.

1: $M_0 \leftarrow M \setminus M_{\text{opt}} \triangleright$ where $M_0$ is the working set of student project pairs in $M$
2: let $L_0$ be a copy of the set of all lecturer clones
3: $L'_0 \leftarrow \emptyset$
4: $M' \leftarrow \emptyset$
5: while $L_0$ is non-empty do
   6:      remove a lecturer clone $l'_k$ from $L_0$
   7:      if $(s_i, l'_k)$ is an edge in $M'_{\text{opt}}$ for some $s_i$ then
      8:         let $p_j$ be the project assigned to $s_i$ in $M_{\text{opt}}$
      9:         if $(s_i', p_j)$ is in $M_0$ for some student $s_i'$ then
         10:            $M' \leftarrow M' \cup \{(s_i', l'_k)\}$
         11:            $M_0 \leftarrow M_0 \setminus \{(s_i', p_j)\}$
      12:            else
      13:               $L'_0 \leftarrow L'_0 \cup \{l'_k\}$
      14:            end if
   15:      else
   16:         $L'_0 \leftarrow L'_0 \cup \{l'_k\}$
   17:      end if
5: end while
18: while $M_0$ is non-empty do
19:      pick some $(s_i, p_j) \in M_0$
20:      let $l'_k$ be some lecturer clone in $L'_0$, where $l_k$ offers $p_j > l'_k$ must exist since there are
21:      $d_k - |M(l_k) \cap M_{\text{opt}}(l_k)|$ clones for $l_k$
22:      $L'_0 \leftarrow L'_0 \setminus \{l'_k\}$
23:      $M' \leftarrow M' \cup \{(s_i, l'_k)\}$
24: end while
25: return $M'$
This example demonstrates the creation of graph $G'$ from graph $G$ and matchings $M$ and $M_{opt}$. Figure 6 shows example instance $I$ of \textsc{spa-st}.

Let $M = \{(s_1, p_1), (s_2, p_2), (s_3, p_3)\}$ and $M_{opt} = \{(s_1, p_1), (s_2, p_2), (s_3, p_2), (s_4, p_3)\}$ be stable matchings in $I$. Clearly, $M_{opt}$ is also a maximum stable matching as all students are assigned. Figure 7a shows the underlying graph $G$ of instance $I$. To create the vertices of $G'$, student vertices are copied, and multiple lecturer cloned vertices are created. For lecturer vertices $l_1$ and $l_2$ in $G$ with capacities of 2, we create $l_1^0, l_1^1$ and $l_2^0$ in $G'$. Using the definition of $M'_{opt}$ above, we obtain the edge set $M'_{opt} = \{(s_2, l_1^0), (s_3, l_1^0), (s_4, l_1^0)\}$. Figure 7b shows a part built $G'$ with all $M'_{opt}$ edges added.

Next $M'$ is calculated using Algorithm 6. A copy of $M \setminus M_{opt}$ is created and denoted $M_0 = \{(s_2, p_2), (s_3, p_3)\}$. $L_0 = \{l_1^0, l_1^1, l_2^0\}$ is a copy of the set of all lecturer cloned vertices, and $L'_0$ is the empty set. We iterate through $L_0$ as follows.

- Lecturer clone $l_1^0$ is removed from $L_0$. Since there is an edge $(s_2, l_1^0) \in M'_{opt}$ and $s_2$ is assigned $p_1$ in $M_{opt}$, but $(s_2, p_1) \notin M_0$ for each student $s_i$, $l_1^0$ is added to $L'_0$;
- Lecturer clone $l_2^0$ is removed from $L_0$. As there is no edge $(s_i, l_2^0) \in M'_{opt}$ for any student $s_i$, $l_2^0$ is added to $L'_0$;
- Next lecturer clone $l_2^0$ is removed from $L_0$. There is an edge $(s_3, l_2^0) \in M'_{opt}$, $s_3$ is assigned $p_2$ in $M_{opt}$ and there is an edge $(s_2, p_2) \in M_0$, hence $(s_2, l_2^0)$ is added to $M'$ and $(s_2, p_2)$ is removed from $M_0$;
- Using the same reasoning when the final lecturer clone $l_2^0$ is removed from $L_0$, $(s_3, l_2^0)$ is also added to $M'$ and $(s_3, p_3)$ is removed from $M_0$.

As $M_0$ is now empty, we do not enter the final while loop on Line 19 of Algorithm 6, therefore $M'$ is now complete. Figure 7c shows the completed mapped graph $G'$ with edge set $M' \cup M'_{opt}$.

**B.4.3 Components in $G'$**

An alternating path in $G'$ is defined as a path that comprises edges in $M_{opt}$ and in $M$ alternately. A path or alternating path is described as even if there are an even number of edges in the path, odd otherwise. Finally, an alternating cycle is a sequence of edges in $M_{opt}$ and $M$ alternately, which forms a cycle.

A component $c$ in $G'$ is defined as any maximal connected subgraph in $G'$. Figure 8 shows the possible component structures that may be found in $G'$ which are described in more detail below. Let $n_{c,l}$ and $n_{c,s}$ denote the maximum number of lecturer clones and students respectively, in some component $c$ of $G'$, and let $n_c = \max\{n_{c,l}, n_{c,s}\}$. Notation for a lecturer clone in component $c$ is defined as $l^{c,r}$ indicating the $r^{th}$ lecturer clone of component $c$. Similarly, $s^{c,r}$ indicates the $r^{th}$ student of component $c$.

**Figure 6** Example instance $I$ of $\textsc{spa-st}$.
Each vertex in $G'$ is incident to at most one $M'$ edge and at most one $M_{opt}'$ edge, meaning every component must be a path or cycle comprising alternating $M'$ and $M_{opt}'$ edges. Therefore the structure of each component must have one of the following forms.

(a) An alternating cycle;
(b) An even length alternating path, with lecturer clone end vertices;
(c) An even length alternating path, with student end vertices;
(d) An odd length alternating path, with end edges in $M$;
(e) An odd length alternating path, with end edges in $M_{opt}$, for $n_c \geq 3$;
(f) An odd length alternating path, with end edges in $M_{opt}$, for $n_c = 2$;
(g) An odd length alternating path, with end edges in $M_{opt}$, for $n_c = 1$;

We wish to show that any stable matching found by Algorithm 1 must be at least 2/3 the size of $M_{opt}$.

B.4.4 Proofs

Proposition 23. Let $(s_i, l^{c,r}) \in M_{opt}'$ be an edge in $G'$ where $(s_i, p_j) \in M_{opt}$. If $l^{c,r}$ is unmatched in $M'$ or if there exists an edge $(s_i, l^{c,r}) \in M'$ where $s_i$ is assigned a project other than $p_j$ in $M$, then $|M_{opt}(p_j)| > |M(p_j)|$, and hence $p_j$ is undersubscribed in $M$.

Proof. Assume for contradiction that $|M_{opt}(p_j)| \leq |M(p_j)|$. During the execution of Algorithm 1, the first while loop iterates over the lecturer clones in $G'$ once. This means that all edges in $M_{opt}'$ corresponding to edges in $M_{opt}(p_j)$ are iterated over.

Let $l^{c,r}$ be the current lecturer clone being iterated over. We know that $(s_i, l^{c,r}) \in M_{opt}'$ where $(s_i, p_j) \in M_{opt}$. If there is an edge $(s_i', p_j) \in M$ not yet mapped then we immediately ‘pair’ on this lecturer clone by adding edge $(s_i, l^{c,r})$ to $G'$. Since $|M_{opt}(p_j)| \leq |M(p_j)|$, there are at least $|M_{opt}(p_j)|$ opportunities for these pairings to occur. Therefore, every $c^{c,r}$ clone corresponding to an edge in $M_{opt}(p_j)$ must be assigned an additional edge in $M'$ which itself
A $\frac{3}{2}$-approximation algorithm for the Student-Project Allocation problem

Figure 8 The possible component structures in $G'$ for a component $c$, where $\gamma = n_c$, the size of the component, and $\mu = \gamma - 1$. $M'$ and $M'_{opt}$ edges are shown in bold and non-bold, respectively.
corresponds to an edge in $M(p_j)$. This provides the required contradiction to the assumption that $t_{c,r}$ is unmatched in $M'$ or $(s_{i'}, t_{c,r}) \in M'$ where $s_{i'}$ is assigned a project other than $p_j$ in $M$. It follows immediately that $p_j$ is undersubscribed in $M$.

**Proposition 24.** Let $(s_i, t_{c,r}) \in M'$ be an edge in $G'$ where $(s_i, p_j) \in M$. If there exists an edge $(s_{i'}, t_{c,r}) \in M'_{\text{opt}}$ where $s_{i'}$ is assigned a project other than $p_j$ in $M_{\text{opt}}$, then $|M(p_j)| > |M_{\text{opt}}(p_j)|$, and hence $p_j$ is undersubscribed in $M_{\text{opt}}$.

**Proof.** Assume for contradiction that there exists an edge $(s_{i'}, t_{c,r}) \in M'_{\text{opt}}$ where $s_{i'}$ is assigned a project other than $p_j$ in $M_{\text{opt}}$ and $|M(p_j)| \leq |M_{\text{opt}}(p_j)|$.

When we iterate over the lecturer clones in the first while loop of Algorithm 6, if there is an edge $(s_{i'}, p_j) \in M$ not yet mapped then we immediately ‘pair’ this lecturer clone. Clearly all pairs of this type in $M$ will be used up since there are only $|M(p_j)|$ opportunities for these pairings to occur. But since an edge $(s_{i'}, t_{c,r}) \in M'$ where $s_{i'}$ is assigned a project $p_j$, this is not the case and we have our required contradiction. Hence $p_j$ is undersubscribed in $M_{\text{opt}}$.

**Proposition 25.** Let $c$ be the component of $G'$ in Figure 8f. Let $s^{c,1}$ be assigned to project $p_j$ in $M_{\text{opt}}$. Then project $p_j$ is fully available in $M$ and, $s^{c,1}$ and $s^{c,2}$ can never have applied to $p_j$ at any point in Algorithm 7’s execution.

**Proof.** Let lecturer clone $l^{c,1}$ correspond to lecturer $l_k$. In $G'$, lecturer clone $l^{c,1}$ is unassigned in $M'$, therefore we know that $l_k$ is undersubscribed in $M$. Since there is no edge in $M'$ incident to $l^{c,1}$, by Proposition 23 we know $p_j$ is undersubscribed in $M$. Project $p_j$ is by definition fully available in $M$ and by Proposition 8 must have been fully available throughout the algorithm’s execution.

Now we prove that neither $s^{c,1}$ nor $s^{c,2}$ can have applied to $p_j$. In Algorithm 3 students can only apply to projects that are not fully available by Proposition 9 hence we only look at the main while loop of Algorithm 1. We consider $s^{c,1}$ first. Assume for contradiction that $s^{c,1}$ applied to $p_j$ at some point during the main while loop of Algorithm 1’s execution. Then $(s^{c,1}, p_j)$ would be added to $M$ as $p_j$ was always fully available. But we know that $(s^{c,1}, p_j)$ is not in the final matching $M$ hence it must have been rejected by $l_k$ at some point. But this can only have happened if $p_j$ was not fully available, which contradicts the fact that $p_j$ is always fully available above. Therefore $s^{c,1}$ can never have applied to $p_j$ at any point. By identical reasoning $s^{c,2}$ can also never have applied to $p_j$.

**Proposition 26.** Suppose that student $s_i$ applied to project $p_j$ in phase 2 of Algorithm 2 and denote this time by $T_0$. Then at time $T_0$, $p_j$ is not fully available and is non-precarious.

**Proof.** Let $l_k$ be the lecturer offering $p_j$. Assume for contradiction that $p_j$ is fully available at $T_0$. Since $s_i$ is applied to $p_j$ in phase 2, $l_k$ must have rejected $s_i$ when $s_i$ was in phase 1. But this can only happen if $p_j$ is not fully available and by Proposition 8 $p_j$ cannot again become fully available.

Assume then that $p_j$ is precarious at $T_0$. Then there must exist a precarious pair $(s_{i'}, p_j)$ in the matching for some student $s_{i'}$. We know from Proposition 10 that when a project is not fully available and non-precarious, it cannot again become precarious. Therefore, when $s_i$ applied in phase 1 to $p_j$, it was either fully available or precarious (or both), and so $(s_i, p_j)$ must have been added to the matching. But at some point before $T_0$, since $s_i$ is applying in phase 2, $(s_i, p_j)$ was removed from the matching. This can only happen when $p_j$ is not fully available and either $(s_i, p_j)$ is precarious or is a worst student in $M(p_j)$ (also a worst assignee of $M(l_k)$).
If \((s_i, p_j)\) is precarious then, once removed, \(s_i\) would again apply to \(p_j\) in phase 1 and must be successfully added for the same reason as before, although this time as a non-precarious pair (since other fully available projects tied with \(p_j\) on \(s_i\)’s list would be applied to by \(s_i\) before \(p_j\)). The removal of non-precarious \((s_i, p_j)\) can only happen because \(p_j\) is non-precarious which contradicts the assumption that \(p_j\) is precarious at \(T_0\) by Proposition 10 since \(p_j\) is also not fully available. Therefore \(p_j\) is non-precarious at \(T_0\).

Therefore \(p_j\) can be neither fully available nor precarious at \(T_0\).

\[\rightarrow\]  
**Proposition 27.** Let \(T_0\) denote the point in Algorithm 7’s execution at the end of the main while loop. If a project \(p_j\) offered by \(l_k\) is full and non-precarious before \(T_0\), then a student \(s_i\) worse than or equal to \(l_k\)’s worst ranked assignees in \(M(p_j)\) cannot subsequently become assigned to \(p_j\).

**Proof.** We know that a student worse than or equal to \(l_k\)’s worst ranked assignees in \(M(p_j)\) cannot become assigned to \(p_j\) before \(T_0\) by Proposition 13. It remains to strengthen this statement to be true for the rest of the algorithm.

After \(T_0\), if there is a blocking pair associated with a student assigned to \(p_j\) then by Propositions 15 and 16 we know that at \(T_0\), \(M(p_j)\) must contain one of the worst students in \(M(l_k)\). Therefore, the only way a worse ranked student could become assigned to \(p_j\) after \(T_0\) is if \(p_j\) is undersubscribed at \(T_0\).

But this would mean that there was some point \(T_1\) before \(T_0\) and after \(p_j\) was full and non-precarious, that \(p_j\) became undersubscribed by the removal of \((s_i, p_j)\) for some student \(s_i\). Denote this point as \(T_2\). The removal of \((s_i, p_j)\) occurring at \(T_2\) can only have happened if \(l_k\) was full, non-precarious (since we are removing a non-precarious pair) and \(l_k\) preferred some other student \(s_i\) to \(s_i\). But by Proposition \(13\) \(l_k\) cannot assign to a worse student after this point. Therefore \(l_k\) cannot subsequently assign to a worse student than \(s_i\) (the worst student in \(M(p_j)\) when \(p_j\) was full).

\[\rightarrow\]  
**Lemma 28.** Let \(M\) be a stable matching found by Algorithm 7 for instance \(I\) of SPA-ST, and let \(M_{opt}\) be a maximum stable matching in \(I\). No component of the type given in Figure 9 can exist in the mapped graph \(G'\).  

**Proof.** Assume for contradiction that there is a component \(c\) of the type shown in Figure 8 in \(G'\). Let \(p_j\) be the project assigned to \(c^{c-1}\) in \(M'_{opt}\) in Figure 8.

We look at the possible configurations in \(G\) that could map to \(c\) in \(G'\). Lecturer clones \(l^{c-1}\) and \(l^{c-2}\) may or may not be the same lecturer in \(G\). It may also be the case that \(s^{c-1}\) and \(s^{c-2}\) are assigned to the same or different projects in \(M'\) and \(M'_{opt}\), respecting the fact that projects may only be offered by one lecturer. Let \(s^{c-1} = s_i\) and \(s^{c-2} = s_{i'}\). Figure 9 shows the possible configurations in \(G\) relating to \(c\) in \(G'\). They are found by noting that all configurations in \(G\) must have: 2 students; 1 or 2 lecturers; between 2 and 3 projects; student \(s_{i'}\) must be unassigned in \(M\); and \(s_i\) must be assigned a project of the same lecturer in \(M_{opt}\) as \(s_i\) is in \(M\). Note that it is not possible for there to be only one project \(p_j\) in the configuration since \(s_i\) would be assigned to \(p_j\) in both \(M\) and \(M'_{opt}\), meaning \((s_i, p_j)\) would not exist in \(M_{opt}\) in \(M'\) or \(M\setminus M_{opt}\) and so neither of the edges from \(s^{c-1}\) would exist in \(G'\), a contradiction.

We now show that none of the subgraphs shown in Figure 9 can occur in a matching \(M\) with respect to \(G\) found using Algorithm 1. Assume for contradiction that one does occur. We consider each type of subgraph separately.

(a) Students \(s_i\) and \(s_{i'}\) are assigned to \(p_j\) and \(p_j'\) in \(M_{opt}\) respectively, and \(s_i\) is assigned to \(p_{j'}\) in \(M\). Lecturer \(l_k\) offers both \(p_j\) and \(p_j'\).
There are three sub-cases to consider.

i. $s_i$ strictly prefers $p_j$ to $p_j'$: If $s_i$ prefers $p_j$ to $M(s_i)$ then $s_i$ must have applied to $p_j$ at least once. But this contradicts Proposition 25.

ii. $p_j$ and $p_j'$ are tied on $s_i$'s preference list: Project $p_j$ is fully available in the finalised matching $M$ by Proposition 25 and has always been fully available by Proposition 8. As there is a fully available project tied with $p_j'$ on $s_i$'s list, once edge $(s_i, p_j')$ is added to $M$, as long as it remains, it must be precarious. Pair $(s_i, p_j')$ cannot be removed at any stage before the end of the main while loop, since doing so would mean $s_i$ would apply to $p_j$ before again applying to $p_j'$ (since $p_j$ is fully available) contradicting Proposition 25. Also Algorithm 3 cannot change the allocations of any precarious lecturer by Proposition 11. Therefore Algorithm 1 must terminate with $(s_i, p_j')$ as a precarious pair.

Student $s_{i'}$ must have applied to $p_j'$ in phase 2 since they are unassigned in the finalised matching $M$. By Proposition 26 at the point of application, $p_j'$ is not fully available and is non-precarious. But by Proposition 10 $p_j'$ cannot subsequently become precarious and so the algorithm will terminate with a non-precarious $p_j'$, contradicting the above.

iii. $s_i$ strictly prefers $p_j'$ to $p_j$: We consider three further sub-cases based on $l_k$'s preference list.

1. $l_k$ strictly prefers $s_i$ to $s_{i'}$: We know that $s_i$ strictly prefers $p_j'$ to $p_j$ and that $l_k$ strictly prefers $s_i$ to $s_{i'}$. But then $(s_i, p_j')$ forms a blocking pair of stable matching $M_{opt}$, a contradiction.
2. $s_i'$ and $s_i$ are tied on $l_k$’s preference list: Project $p_j$ is fully available in $M$ by Proposition 25 and has always been fully available by Proposition 8. Student $s_i$ must have been assigned to $p_j$ in phase 1, otherwise $s_i$ would have applied to $p_j$, contradicting Proposition 25. Student $s_i'$ must have applied to $p_j$ in phase 2 since $s_i'$ is unassigned. By Proposition 26, at time $T_0$, $p_j$ is not fully available and is non-precarious. Regardless of whether $(s_i, p_j)$ exists in the matching at time $T_0$, we know from time $T_0$, $(s_i, p_j)$ cannot be removed from $M$, otherwise $s_i$ would remove $p_j$ from their preference list (as $p_j$ is not precarious) contradicting the fact that $s_i$ assigned to $p_j$ in phase 1.

We consider the following two possibilities.

- $s_i'$ applied to $p_j$ in phase 2 before pair $(s_i, p_j)$ was added: Assume $s_i'$ is unsuccessful in its application at $T_0$. From above we know that $p_j$ remains non-precarious from $T_0$ onwards.
  
  - If $p_j$ is undersubscribed at time $T_0$, then $l_k$ cannot be precarious (as $s_i'$ was rejected) and so $l_k$ does not meta-prefer $s_i'$ to its worst assignee $s_w \in M(l_k)$. Lecturer $l_k$ must be full at this point since $p_j$ is not fully available and is undersubscribed. As $l_k$ is full and must remain non-precarious by Proposition 10 after $T_0$, it is only possible for $l_k$ to improve their allocations, by Proposition 13. Since $l_k$ meta-prefers $s_j$ to $s_i$ ($s_i'$ is in phase 2), when $s_i$ applies to $p_j$, $s_i$ must also be rejected. Project $p_j$ is not precarious and so $s_i$ must remove $p_j$ from their list, contradicting the fact that $s_i$ must be assigned to $p_j$ in phase 1.

- If $p_j$ is full at $T_0$ then we know $l_k$ does not meta-prefer $s_i'$ to its worst assignee $s_w \in M(p_j)$. By Proposition 13, $p_j$ cannot accept worse assignments than are currently in $M(p_j)$ until after the main while loop, therefore when $s_i$ applies to $p_j$ before the main while loop as in the previous case they must also be rejected, a contradiction as above.

Assume therefore that $s_i'$ is successful in their application at $T_0$. Pair $(s_i', p_j)$ must be removed at some point after $T_0$ since $(s_i', p_j) \notin M$. Denote the time $(s_i', p_j)$ is removed as $T_1$. The removal at $T_1$ must have occurred before the end of the main while loop since otherwise $s_i'$ would be assigned to some project in the finalised matching $M$, a contradiction (the same students are assigned when removing blocking pairs of type (3bi)). We know that $(s_i, p_j)$ was added either after $T_0$ and before $T_1$ or after $T_1$. Once added $(s_i, p_j)$ cannot be removed from above.

- Assume $(s_i, p_j)$ was added before $T_1$. At $T_1$ (before the end of the main while loop) pair $(s_i', p_j)$ is removed. This must either be because $p_j$ is undersubscribed and $l_k$ is full, or because $p_j$ is full.
  
  * If the former then $l_k$ is full and cannot be precarious since we are removing a non-precarious pair $(p_j)$ is non-precarious after $T_0$). But this removal can only happen if $s_i'$ is the worst student assigned to $l_k$ at $T_1$. But by the definition of a worst assignee $s_i$ (being in phase 1) would be removed before $s_i'$. Therefore, $(s_i, p_j)$ must have already been removed from the matching, a contradiction to the fact that $(s_i, p_j)$ cannot be removed.

* Using similar reasoning, if $p_j$ is full at $T_1$, then (as we are removing a non-precarious pair), $s_i'$ must be the worst student assigned in $M(p_j)$.
But this would mean \((s_i, p_{j'})\) had already been removed, a contradiction.

Assume \((s_i, p_{j'})\) was added after \(T_1\). Again we consider 2 sub-cases.

* If \(p_{j'}\) was undersubscribed at time \(T_1\), then \(l_k\) must have been full and must be non-precarious since non-precarious pair \((s'_{i'}, p_{j'})\) was removed. This pair was removed as \(s'_{i'}\) was a worst student in \(l_k\) at \(T_1\). By Proposition 10, \(l_k\) remains non-precarious from this point onwards and therefore by Proposition 13, \(l_k\) can only improve their allocations from time \(T_1\). Therefore, as \(l_k\) meta-prefers \(s'_{i'}\) to \(s_i\) (as \(s'_{i'}\) is in phase 2), it must be that \(s_i\); applying after \(T_1\), will be rejected. This would result in the removal of \(p_{j'}\) from \(s_i\)’s list contradicting the fact that \(s_i\) assigned to \(p_{j'}\) in phase 1.

* If \(p_{j'}\) is full at \(T_1\) then using similar reasoning to above, we know that at \(T_1\), \(s'_{i'}\) is a worst assignee in \(M(p_{j'})\). Since, at \(T_1\), \(p_{j'}\) is full and non-precarious, and remains non-precarious, by Proposition 13, \(p_{j'}\) cannot subsequently accept a worse assignee until the end of the main while loop.

Therefore \(s_i\) will be rejected on application, a contradiction as above.

\(- s'_{i'}\) applied to \(p_{j'}\) in phase 2 after pair \((s_i, p_{j'})\) was added: Since \(l_k\) meta-prefers \(s'_{i'}\) to \(s_i\), \((s'_{i'}, p_{j'})\) must be added to the matching with some student other than \(s'_{i'}\) being removed. But now we are in the same position as before where \((s'_{i'}, p_{j'})\) must be removed from \(M\), but this can only happen if \((s_i, p_{j'})\) is removed first, a contradiction.

3. \(l_k\) strictly prefers \(s'_{i'}\) to \(s_i\): Since \(s'_{i'}\) is unassigned in \(M\), \((s'_{i'}, p_{j'})\) is a blocking pair of stable matching \(M\), a contradiction.

(b) Students \(s_i\) and \(s'_{i'}\) are both assigned to the same project \(p_j\) in \(M_{\text{opt}}\) and \(s_i\) is assigned to project \(p_{j'}\) in \(M\). Both \(p_j\) and \(p_{j'}\) are offered by lecturer \(l_k\). Since \(s'_{i'}\) is unassigned in \(M\), we know that \(s'_{i'}\) has to have applied to \(p_j\) during the algorithm’s execution, but this directly contradicts Proposition 25.

(c) Students \(s_i\) and \(s'_{i'}\) are assigned to \(p_j\) and \(p_{j'}\) in \(M_{\text{opt}}\) and \(s_i\) is assigned to \(p_{j'}\) in \(M\). Lecturer \(l_k\) offers \(p_j\), \(p_{j'}\) and \(p_{j''}\). By Proposition 25, \(p_{j''}\) is undersubscribed in \(M\) since lecturer clone \(l'\) in Figure 9 has two edges corresponding to different projects in \(M\) and \(M_{\text{opt}}\). We know that \(l_k\) is undersubscribed in \(M\) by Proposition 25 and so \(p_{j''}\) must also be fully available. By Proposition 8, we know that \(p_{j''}\) has always been fully available during the algorithm’s execution. Since \(s'_{i'}\) is unassigned in \(M\) they must have applied to \(p_{j''}\) during the course of the algorithm. But as \(p_{j''}\) has always been fully available this must have been accepted. As we end up with \(s'_{i'}\) being unassigned it must also be the case that \(l_k\) rejects pair \((s'_{i'}, p_{j''})\) but we know that \(p_{j''}\) is always fully available and so this cannot have happened, a contradiction.

(d) Students \(s_i\) and \(s'_{i'}\) are assigned to \(p_j\) and \(p_{j'}\) in \(M_{\text{opt}}\) and \(s_i\) is assigned to \(p_{j'}\) in \(M\). Lecturers \(l_k\) and \(l_{k'}\) offer projects \(p_j\) and \(p_{j'}\), respectively. Identical arguments to those found in Case 3 can be used to show a contradiction. Similarly, identical arguments to those found in Case 4 can also be used to show a contradiction, but exchanging \(l_k\) for \(l_{k'}\).

(e) Students \(s_i\) and \(s'_{i'}\) are both assigned to project \(p_j\) in \(M_{\text{opt}}\) and \(s_i\) is assigned to project \(p_{j'}\) in \(M\). Projects \(p_j\) and \(p_{j'}\) are offered by lecturers \(l_k\) and \(l_{k'}\) respectively. Using identical arguments to Case 3 as student \(s'_{i'}\) is unassigned in \(M\), they must have applied to \(p_j\) during the algorithm’s execution, but this contradicts Proposition 25.
Students $s_i$ and $s_{i'}$ are assigned to $p_j$ and $p_{j'}$, respectively, in $M_{opt}$ and $s_i$ is assigned to $p_{j'}$ in $M$. Lecturer $l_k$ offers project $p_{j'}$ whereas $l_k$ offers projects $p_j$ and $p_{j''}$. We consider the following 3 sub-cases.

i. $s_i$ strictly prefers $p_j$ to $p_{j'}$: Identical arguments to those found in Case 1 can be used to show a contradiction.

ii. $p_j$ and $p_{j'}$ are tied on $s_i$’s preference list: Using similar reasoning to Case 1i, we know that once edge $(s_i, p_{j'})$ is added to $M$ it is, and remains, a precarious pair and cannot be removed at any stage. Also $p_{j'}$ must have been fully available on application by $s_i$ otherwise fully available $p_j$ at the same rank would have been applied to by $s_i$. Therefore $p_{j'}$ is either fully available or precarious throughout the algorithm’s execution. Student $s_i$ is not assigned in $M$ and so must have applied to $p_{j'}$ whilst they were in phase 2. Let this time of application be denoted $T_0$. By Proposition 26, $p_{j''}$ cannot be precarious at $T_0$.

- If $p_{j'}$ is fully available at $T_0$ then $l_{k'}$ is undersubscribed and so $s_{i'}$ would only be rejected if $p_{j''}$ was not precarious, full and $s_{i'}$ was not meta-preferred by $l_{k'}$ to an in $M(p_{j''})$.

- If $p_{j'}$ is precarious at $T_0$ then $l_{k'}$ is also precarious and therefore $s_{i'}$ would again only be rejected if $p_{j''}$ was not precarious, full and $s_{i'}$ was not meta-preferred by $l_{k'}$ to an student in $M(p_{j''})$.

Therefore we have the following two cases.

1. If $s_{i'}$ was rejected then it must be because $p_{j''}$ was not precarious, full and $s_{i'}$ was not meta-preferred by $l_{k'}$ to any student in $M(p_{j''})$, by above. But similar to Case 1 we can say that by Proposition 23, $p_{j''}$ is undersubscribed in the finalised matching $M$. Therefore, at least one non-precarious pair involved with $p_{j''}$ must be removed (without a pair involving $p_{j''}$ immediately replacing it) before the end of the algorithm. Denote this point in the algorithm’s execution as $T_1$ and the removed pair $(s_{i''}, p_{j''})$ for some student $s_{i''}$. Note $T_1$ may either be before or after the end of the main while loop. This type of removal can only happen when $l_{k'}$ is full (this is clear before the main while loop, and is true after the main while loop by Proposition 7), and once a lecturer is full they remain full (since any pair deletion involving a project of $l_{k'}$ can only occur with a pair addition involving a project of $l_{k'}$). But $(s_i, p_{j'})$ was assigned when $p_{j''}$ was fully available and so $(s_i, p_{j'})$ was assigned before $T_1$. If $T_1$ occurs after the end of the main while loop then $l_{k'}$ must be non-precarious at $T_1$ by Proposition 11. But this means precarious pair $(s_i, p_{j'})$ must have been removed before $T_1$, a contradiction to the fact that $(s_i, p_{j'})$ can never be removed. Therefore the removal of $(s_{i''}, p_{j''})$ at $T_1$ must have occurred before the end of the main while loop. But, since $(s_i, p_{j'})$ is precarious, it would be removed before non-precarious $(s_{i''}, p_{j''})$, a contradiction.

2. If $s_{i'}$ was accepted then pair $(s_{i''}, p_{j''})$ would need to be removed before the algorithm terminated (since $(s_{i''}, p_{j''}) \notin M$). We know from before that $p_{j''}$ is non-precarious at the point of application and further that pair $(s_{i''}, p_{j''})$ must remain non-precarious by definition since $s_{i'}$ applied in phase 2. Therefore, we need to remove non-precarious pair $(s_{i''}, p_{j''})$ from the matching which can only happen if either $p_{j''}$ is full or $l_{k'}$ is full. Firstly assume that $p_{j''}$ is full and pair $(s_{i''}, p_{j''})$ is replaced with a meta-preferred student assigned to $p_{j''}$ ($p_{j''}$ must be non-precarious since we are removing a non-precarious pair). Since $p_{j''}$ needs to be undersubscribed in the finalised matching $M$ we are in the same position and
iii. $s_i$ strictly prefers $p_j$ to $p_j'$: We now consider three sub-cases based on $l_{k'}$’s preference list.

1. $l_{k'}$ strictly prefers $s_i$ to $s_{i'}$: We know that $p_j'$ is undersubscribed in $M_{\text{opt}}$ by Proposition 24. Therefore as $l_{k'}$ strictly prefers $s_i$ to $s_{i'}$, and $s_i$ strictly prefers $p_j'$ to $p_j$, $(s_i, p_j')$ is a blocking pair of stable $M_{\text{opt}}$, a contradiction.

2. $s_{i'}$ and $s_i$ are tied on $l_{k'}$’s preference list: For this initial paragraph we use some similar reasoning to Case 11iiia(iii)2. Student $s_i$ must have assigned to $p_j'$ in phase 1, otherwise $s_i$ would have applied to $p_j$ a contradiction to Proposition 25. Unlike Case 11iib(iii)2, $p_j'$ may be precarious at this point of application. Also, student $s_{i'}$, not being assigned in $M$, must have applied to $p_j'$ whilst in phase 2. Denote the point at which $s_{i'}$ applies to $p_j'$ in phase 2 as $T_0$. At $T_0$ we know that $p_j'$ is not fully available and not precarious by Proposition 26 and that $p_j'$ remains non-precarious from this point onwards by Proposition 10.

We look at two possibilities:

A. $s_{i'}$ was rejected at $T_0$. There would be two possible reasons for the rejection.

Firstly, that $l_{k'}$ is non-precarious, full and $l_{k'}$ does not meta-prefer $s_{i'}$ to any student in $M(l_{k'})$. Secondly, that $p_j'$ is non-precarious, full and $l_{k'}$ does not meta-prefer $s_{i'}$ to any student in $M(p_j')$. We can rule out the first option as follows. If $l_{k'}$ is non-precarious and full at $T_0$ then by Proposition 13, $l_{k'}$ cannot accept a worse student than currently exists in $M(l_{k'})$ for the reminder of the algorithm. Since $s_{i'}$ was rejected we can conclude that no worse student than $s_{i'}$ can exist in $M(l_{k'})$ at $T_0$ and cannot exist in $M(l_{k'})$ from $T_0$ onwards. But $s_{i'}$ being in phase 2 is meta-preferred to $s_i$ in phase 1. This contradicts the fact that $(s_i, p_j') \in M$. Therefore, $s_{i'}$ was rejected because $p_j'$ is non-precarious, full and $l_{k'}$ does not meta-prefer $s_{i'}$ to any student in $M(p_j')$. By Proposition 27, $l_{k'}$ cannot accept a student to project $p_j'$ that is worse or equal to a worst student existing in $M(p_j')$ for the remainder of the algorithm.

Using a similar strategy to Case 11iiia, we know that $p_j''$ is undersubscribed in the finalised matching $M$ by Proposition 23, therefore before the algorithm terminates a pair $(s_{i''}, p_j'')$ involving $p_j''$ must be removed without being immediately replaced with another pair involving $p_j''$. Denote the first such occurrence as happening at time $T_1$, where $T_1$ occurs after $T_0$. Note that $T_1$ may be either before or after the end of the main while loop.

Assume $T_1$ occurs before the end of the main while loop. We know any removal of the type occurring at $T_1$ must be due to $l_k$ being full. By Proposition 13, we know $l_{k'}$ cannot be subsequently assigned in $M$ a worse student than exists in the matching at $T_1$. Using the same proposition we know that $p_j''$ cannot be assigned in $M$ a worse student until the end of the main while loop than exists in the matching at $T_0$. Since a pair involving $p_j''$ was removed at $T_1$, a worst assignee in $M(l_{k'})$ at $T_1$ can be no worse than a worst assignee in $M(p_j'')$ at $T_0$. Finally, this means that no student assigned to $l_{k'}$ from $T_1$ onwards can be worse than $s_{i'}$, rejected at $T_0$, but $s_i$ being in phase 1 is worse than $s_{i'}$ in phase 2 according to $l_{k'}$, a contradiction to the fact that $(s_i, p_j') \in M$. 
Assume therefore that $T_1$ occurs after the end of the main while loop. Then $s_i$ has to be assigned to $p_j$ at this point. Since $p_j$ becomes undersubscribed at $T_1$, pair $(s_i, p_j)$ must be a blocking pair of type (3bi). By Propositions 15 and 16, $s_i$ must be a worst student in $M(l_k)$ and therefore $M(p_j)$. But, we also know that at $T_0$ when $s_i$ was rejected, $l_k$ could not subsequently accept a student to project $p_j$ that is worse or equal to a worst student existing in $M(p_j)$. Therefore, $s_i$ can exist in the mapped graph $G$. Also, by Proposition 8, $p_j$ must also be undersubscribed in $M$. Therefore, $p_j$ is available at the end of the algorithm’s execution and must always have been fully available by Proposition 8.

Student $s_i$ is unassigned in $M$ and so we know that $s_i$ has to have applied to $p_j$ during the algorithm’s execution. Since $p_j$ has always been fully available, this had to have been

**Lemma 29.** Let $M$ be a stable matching found by Algorithm 7 for instance $I$ of SPA-ST and let $M_{opt}$ be a maximum stable matching in $I$. No component of the type given in Figure 3 can exist in the mapped graph $G'$. ▲

**Proof.** Let $p_j$ be the project that student $s_i$ is assigned to in $M_{opt}$ and let lecturer clone $l_k$ correspond to lecturer $l_k$ in $G$. $l_k$ must be undersubscribed in $M$ as there is a lecturer clone $l_k$ unassigned in $G'$. Also, by Proposition 23, $p_j$ must also be undersubscribed in $M$. Therefore, $p_j$ is available at the end of the algorithm’s execution and must always have been fully available by Proposition 8.

Student $s_i$ is unassigned in $M$ and so we know that $s_i$ has to have applied to $p_j$ during the algorithm’s execution. Since $p_j$ has always been fully available, this had to have been
Let $M$ be a stable matching found by Algorithm $1$ for instance $I$ of \textsc{spa-st} and let $\mathcal{M}_{opt}$ be a maximum stable matching in $I$. Then matching $|M| \geq \frac{2}{3}|\mathcal{M}_{opt}|$. 

\textbf{Proof.} Let $G'$ be the mapped graph constructed from the underlying graph of the instance $G$. Components in $G'$ may only exist in the forms shown in Figure $8$. Therefore we need only show that no component in $G'$ can exist where the number of $M'$ edges is less than $2/3$ of the number of $\mathcal{M}_{opt}'$ edges. We run through each component of Figure $8$ in turn. Let the current component be denoted $c$, where $M'(c)$ and $\mathcal{M}_{opt}'(c)$ denote the set of edges in $M'$ and $\mathcal{M}_{opt}'$ involved in $c$, respectively.

For Case $8\text{a}$, an alternating cycle, and Cases $8\text{d}$ and $8\text{e}$ alternating paths of even length, it is clear that $|M'(c)| = |\mathcal{M}_{opt}'(c)|$. Case $8\text{b}$ involves an odd length alternating path with end edges in $M'$. It must be the case therefore that $|M'(c)| > |\mathcal{M}_{opt}'(c)|$ for components of this type. Case $8\text{c}$ shows an odd length alternating path with end edges in $\mathcal{M}_{opt}'$, but for path sizes greater than 5. Therefore, $|M'(c)| \geq \frac{2}{3}|\mathcal{M}_{opt}'(c)|$ as required. Neither Case $8\text{f}$ nor $8\text{g}$ can exist in $G'$ by Lemmas $28$ and $29$ respectively.

Hence it is not possible for the mapped graph $G'$ to contain components in which $|M'(c)| < \frac{2}{3}|\mathcal{M}_{opt}'(c)|$. Algorithm $1$ is therefore $3/2$-approximating to a maximum stable matching in $I$. 

\section{Lower bound for Algorithm $1$}

Figure $10$ shows instance $I$ of \textsc{spa-st}. A maximum stable matching $M'$ in $I$ is given by $M' = \{(s_1,p_2), (s_2,p_3), (s_3,p_1)\}$. The only possible blocking pairs for this matching are $(s_3,p_3)$ and $(s_3,p_2)$. Neither pair can be a blocking pair since $l_2$ prefers both of their current assignees to $s_3$.

A trace is given as Table $7$ which shows the execution run of Algorithm $1$ over instance $I$. The algorithm outputs stable matching $M = \{(s_1,p_3), (s_2,p_2)\}$. The possible blocking pairs of this matching are $(s_2,p_3)$ and $(s_3,p_3)$. Neither can be a blocking pair since $l_2$ prefers $s_1$ to both $s_2$ and $s_3$.

Therefore, Algorithm $1$ has found a stable matching that is exactly $\frac{2}{3}$ the size of the maximum stable matching, thus algorithm cannot guarantee a better bound than $\frac{2}{3}$. 

\section{IP model proofs}

\textbf{Lemma 2.} Let $I$ be an instance of \textsc{spa-st} and let $M$ be a matching in $I$. Then $M$ is stable if and only if the following condition, referred to as condition $(*):$ For each
A stable matching in $I$ corresponds to a feasible matching in $J$ and vice versa.

**Theorem 31.** Given an instance $I$ of $\text{SPA-ST}$, let $J$ be the IP model as defined in Figure 7. A stable matching in $I$ corresponds to a feasible matching in $J$ and vice versa.

**Proof.** Assume instance $I$ of $\text{SPA-ST}$ contains a stable matching $M$. We construct a feasible solution to $J$ involving the variables $x$, $\alpha$ and $\beta$ as follows.

The variables $x$, $\alpha$ and $\beta$ are constructed as follows. For each student $s_i \in S$ and for each project $p_j \in P$, if $s_i$ is assigned to $p_j$ in $M$ then we set variable $x_{ij} = 1$, otherwise $x_{ij} = 0$. Let lecturer $l_k$ be the proposer of project $p_j$. Let variable $\alpha_{ij} = 1$ if the following two conditions hold: i) student $s_i$ is not assigned to lecturer $l_k$, and ii) lecturer $l_k$ is full and prefers their worst ranked assignee to $s_i$, or is indifferent between them. Else let $\alpha_{ij} = 0$. Let variable $\beta_{ij} = 1$ if the following two conditions hold: i) student $s_i$ is not assigned to project
$p_j$, and $ij$ project $p_j$ is full, and $l_k$ prefers $p_j$’s worst assignee to $s_i$, or is indifferent between them. Else, set $\beta_{ij} = 0$.

Now it must be shown that all constraints described in Figure 1 are satisfied.

1. **Constraints 1 - 4.** It is clear by the construction of $J$ that Constraints 1-4 are satisfied.

2. **Constraint 5.** Recall $S_{ij} = \{p_r \in P : \text{rank}(s_i, p_r) \leq \text{rank}(s_i, p_j)\}$ is the set of projects ranked at least as highly as $p_j$ in $s_i$’s preference list. Let $\gamma_{ij} = 1 - \sum_{p_r \in S_{ij}} x_{ir}$. We must show that whenever $\gamma_{ij} = 1$, $\alpha_{ij} + \beta_{ij} \geq 1$. Assume $\gamma_{ij} = 1$, that is $s_i$ is unassigned or would prefer to be assigned to $p_j$ than to $M(p_j)$. As $M$ is stable we know that condition (*) of Lemma 2 is satisfied. Therefore $\alpha_{ij} + \beta_{ij} \geq 1$ by construction. This directly satisfies Constraint 5.

3. **Constraint 6.** Recall that for student $s_i$ and lecturer $l_k$, $T_{ik}$ is the set of students ranked at least as highly as student $s_i$ in lecturer $l_k$’s preference list, not including $s_i$. Assume $\alpha_{ij} = 1$. Then, by definition, we know that lecturer $l_k$ is full and prefers their worst ranked assignee to $s_i$, or is indifferent between them. Therefore, the LHS of the inequality must equal $d_k$ and so this constraint is satisfied.

4. **Constraint 7.** Recall that $T_{ijk}$ is the set of students ranked at least as highly as student $s_i$ in lecturer $l_k$’s preference list, such that the project $p_j$ is acceptable to each student. Similar to above, assume $\beta_{ij} = 1$. Then, by definition, we know that project $p_j$ is full and $l_k$ prefers their worst ranked assignee in $M(p_j)$ to $s_i$, or is indifferent between them. Therefore, the LHS of the inequality must equal $c_j$ and so this constraint is also satisfied.

We have shown that the assignment of values to $x$, $\alpha$ and $\beta$ satisfy all the constraints in $J$, thus if there is a stable matching $M$ in $I$, then there is a feasible solution of $J$.

Conversely, we now show that a feasible solution of $J$ corresponds to a stable matching $M$ in $I$. Let $x$, $\alpha$ and $\beta$ be a feasible solution of $J$. For each $x_{ij}$ variable in $J$, if $x_{ij} = 1$ then add pair $(s_i, p_j)$ to $M$ in $I$. It is now shown that this assignment of students to projects satisfies the definition of a stable matching $M$ in $I$.

1. The following constraints are clearly satisfied by Constraints 1-4:
   - A student $s_i$ may be assigned to a maximum of 1 project.
   - A student $s_i$ may only be assigned to a project that they find acceptable.
   - The number of students assigned to project $p_j$ is less than or equal to $c_j$.
   - The number of students assigned to projects offered by lecturer $l_k$ is less than or equal to $d_k$.

2. **$M$ is stable.** Assume for contradiction that there exists a blocking pair $(s_i, p_j) \in M$. Then by Lemma 2 neither of the sub-conditions of condition (*) can be true. Both of these sub-conditions being false imply that, as Constraints 6 and 7 must be satisfied, $\alpha_{ij} = 0$ and $\beta_{ij} = 0$.

Recall $\gamma_{ij} = 1 - \sum_{p_r \in S_{ij}} x_{ir}$. $\sum_{p_r \in S_{ij}} x_{ir}$ is the number of projects that student $s_i$ is assigned to at a higher or equal ranking than $p_j$ in $s_i$’s preference list (including $p_j$). Since $(s_i, p_j)$ is a blocking pair, then it must be the case that $\sum_{p_r \in S_{ij}} x_{ir} = 0$. But this forces $\gamma_{ij} = 1$, and we know that $\alpha_{ij}$ and $\beta_{ij}$ are equal to 0 so Constraint 5 is contradicted.

We have shown that if there is a feasible solution of $x$, $\alpha$ and $\beta$ of $J$, then there is a stable matching $M$ in $I$. This completes the proof. ▽
Corollary 32. \( \text{Given an instance } I \text{ of } \text{SPA-ST}, \text{ let } J \text{ be the IP model as defined in Figure 1. A maximum stable matching in } I \text{ corresponds to an optimal solution in } J \text{ and vice versa.} \)

Proof. Assume \( M \) is a maximum stable matching in \( I \). Let \( f = \langle x, \alpha, \beta \rangle \) be the solution in \( J \) constructed according to the description in Theorem 31. We must show that \( f \) forms an optimal solution of \( J \). Firstly, since \( M \) is stable, we know by Theorem 31 that \( f \) is a feasible solution of \( J \). Suppose for contradiction that \( f \) is not optimal. Then there is some solution \( g = \langle x', \alpha', \beta' \rangle \) of \( J \) in which \( \text{obj}(g) > \text{obj}(f) \), where \( \text{obj}(f') \) gives the objective value of \( f' \). But by construction, \( g \) would translate into a stable matching \( M' \) such that \( |M'| = \text{obj}(g) > \text{obj}(f) = |M| \) in \( I \) contradicting the fact that \( M \) is maximum.

Conversely, assume \( f = \langle x, \alpha, \beta \rangle \) is an optimal solution in \( J \), and let \( M \) be the stable matching in \( I \) constructed according to the description in Theorem 31. Suppose for contradiction that there is some stable matching \( M' \) in \( I \) such that \( |M'| > |M| \). Then by construction, there must be some corresponding solution \( g = \langle x', \alpha', \beta' \rangle \) of \( J \) such that \( \text{obj}(g) = |M'| > |M| = \text{obj}(f) \), giving the required contradiction. \( \square \)

\(^3\) Corollary 32 is identical to Theorem 3 in the main body of the paper.