Operations Research Letters 49 (2021) 655–662

Contents lists available at ScienceDirect

Operations Research Letters

www.elsevier.com/locate/orl

(Un)stable matchings with blocking costs

Yuri Faenza [a], Ioannis Mourtos [b,*], Michalis Samaris [b], Jay Sethuraman [a]

[a] IEOR Department, Columbia University, NY, USA
[b] Management Science and Technology Department, Athens University of Economics and Business, Athens, Greece

ARTICLE INFO

Article history:
Received 20 August 2020
Received in revised form 17 June 2021
Accepted 6 July 2021
Available online 13 July 2021

Keywords:
Stable matching
Blocking edges
Rotation poset

ABSTRACT

We introduce and study weighted bipartite matching problems under strict preferences where blocking edges can be paid for, thus imposing costs rather than constraints as in classical models. We focus on the setting in which the weight of an edge represents the benefit from including it in the matching and/or the cost if it is a blocking edge. We show that this setting encompasses interesting special cases that remain polynomially-solvable, although it becomes APX-hard even in a quite restricted case.

1. Introduction

Gale and Shapley's concept of stability [7] has been widely studied in theory and applied in the practice of two-sided matching markets. In the classical model, we are given as input a bipartite graph $G(V,E)$ whose nodes represent the *agents*, together with a collection $\{>_x\}_{x\in V}$, where for $x \in V$, $>_x$ is a strict preference order imposed by agent x on her neighbors in G. The pair $(G,>)$ is called a *preference system*, and we represent the two sides of the bipartition of G by M (the set of men) and W (the set of women). We use the convention that $y >_x y'$ means that x prefers y to y'. A matching of G is *stable* in $(G,>)$ if it has no *blocking edge*, i.e., a pair of agents preferring each other to their partners in the matching, where being unmatched is less preferable than being matched to any of the partners in an agent's list.

In some contexts, however, we wish to relax stability. For instance, in school choice applications, we may accept some degree of instability if doing so leads to a substantial improvement in the quality of the matching for the students. One option is therefore to apply algorithms like *Top Trading Cycle* (see Abdulkadiroğlu and Sönmez [1]) or relax the stability condition to *legality* (Ehlers and Morrill [5], see also Faenza and Zhang [6]). However, these relaxations do not control the number of blocking edges that are introduced.

A different approach is taken by Biró et al. [4], who study the problem of computing a matching of maximum cardinality that has the smallest number of blocking edges (MMSBP). This formulation leads unfortunately to strong intractability results: Hamada et al. [9] showed that the problem is NP-hard to approximate better than a factor $n^{1-\delta}$ for each $\delta \in (0,1)$, where n is the size of one side of the agents, even when all preference lists have length at most 3. Another related concept, introduced by Askalidis et al., is that of a *socially stable matching* [2], in which some blocking edges are allowed. This models a common situation in large markets where many agents do not know each other (or each other's preferences), so they may not be aware that they form a blocking edge. Finding a largest socially stable matching is also NP-Hard, but approximable within a factor of 3/2 [2].

These results from the literature seem to suggest that relaxing stability while keeping some control on the blocking edges leads to models that are computationally intractable. In this paper, we show this does not have to be the case: we propose computationally tractable problems that relax stability while at the same time restricting the cost incurred by the presence of blocking edges. We take the point of view of a central planner who chooses the output matching μ so as to maximize the revenue produced by edges in μ, minus the cost incurred by edges blocking μ. In the following, for a matching μ in a graph $G(V,E)$, we let B_μ be the edges of G blocking μ. The first natural problem to consider is the following `General Unstable Matching` problem. As is customary, for a set S we let $p(S) := \sum_{e\in S} p(e)$.

Given: A preference system $(G(V,E),>)$, $p,c : E \to \mathbb{Z}$.
Find: A matching μ of G maximizing $p(\mu) - c(B_\mu)$.

(Note that we are allowing both p and c to have negative values here.) This problem is however already NP-Hard when $p(e) = \alpha$, $c(e) = \beta$ for every e, since each instance of MMSBP can be formulated as an instance of the `General Unstable Matching`

* Corresponding author.
E-mail addresses: yf2414@columbia.edu (Y. Faenza), mourtos@aueb.gr (I. Mourtos), michsam@aueb.gr (M. Samaris), jay@ieor.columbia.edu (J. Sethuraman).

https://doi.org/10.1016/j.orl.2021.07.005
0167-6377/

problem by letting $\alpha \gg \beta > 0$. We therefore investigate simplified variations.

We assume first that there is unique nonnegative weight $p(e)$ associated with each edge e, representing *both* the revenue obtained if $e \in \mu$ *and* the cost incurred if e blocks μ. Hence, the value of a matching μ is $f(\mu) := p(\mu) - p(B_\mu)$.

Within this framework, we consider three different problems. First, we restrict the image of p to $\{0, 1\}$, so as to obtain the `Binary Unstable Matching` problem. It includes the special case in which $p(e) = 1$ for all $e \in E$.

Given: A preference system $(G(V, E), >)$ and a function $p : E \to \{0, 1\}$.
Find: A matching μ of $(G(V, E), >)$ maximizing $f(\mu)$.

In our second problem, we interpret $p(e)$ as reflecting the "importance" of edge e in the endpoints' preference lists. Formally, we say that $p : E \to \mathbb{Z}_{\geq 0}$ is *preference-concordant* if, for each $xy, xy' \in E$, $y >_x y'$ whenever $p(xy) > p(xy')$ (in Section 2.2 we argue that this property cannot be assumed without loss of generality). The `Preference Concordant Unstable Matching` problem is thus defined as follows.

Given: A preference system $(G(V, E), >)$ and a preference-concordant function $p : E \to \mathbb{Z}_{\geq 0}$.
Find: A matching μ of $(G(V, E), >)$ maximizing $f(\mu)$.

Theorem 1 collects our first two results.

Theorem 1. *The* `Binary Unstable Matching` *and* `Preference Concordant` *problems can be solved in polynomial time.*

In light of these results, it is tempting to relax the assumptions on the function p. We show, however, that even a mild relaxation of the `Binary Unstable Matching` problem leads to a problem that is hard to approximate. Consider the following β-`Binary Unstable Matching` problem.

Given: A preference system $(G(V, E), >)$, $\beta \in (0, 1)$ and a function $p : E \to \{\beta, 1\}$.
Find: A matching μ of $(G(V, E), >)$ maximizing $f(\mu)$.

Theorem 2. *The* β-`Binary Unstable Matching` *problem is* `APX-Hard`*.*

Finally, we consider an alternative way of restricting the `General Unstable Matching` problem, where each edge satisfies either $c(e) = +\infty, p(e) \geq 0$, or $p(e) = -\infty, c(e) \geq 0$. Edges of G can therefore be partitioned into "blue" – that can be part of the output matching μ, but are not allowed to block it – and "red" – that can block μ at a cost, but cannot be part of it. We call it therefore the `Red-Blue Unstable Matching` problem and define it using only the weight function p.

Given: A preference system $(G(V, E), >)$, a function $p : E \to \mathbb{Z}_{\geq 0}$ and a set $F \subseteq E$ ("red edges").
Find: A stable matching μ of $(G(V, E \setminus F), >)$ maximizing $f(\mu)$.

Theorem 3. *The* `Red-Blue Unstable Matching` *problem can be solved in polynomial time.*

We prove Theorems 1, 2, and 3 in Sections 2, 3, and 4, respectively. Throughout the paper, we rely on the following notation: for matchings μ, S of a graph $G(V, E)$, we let $B_\mu^S := S \cap B_\mu$ and $V(S)$ be the set of nodes matched by S. For $v \in V$, we let $\mu(v)$ be the partner of v in μ (or $\emptyset$ if it does not exist) and say that v *prefers* S *to* μ if $S(v) >_v \mu(v)$. For $n \in \mathbb{N}$, we write $[n] := \{1, \ldots, n\}$ and $[n]_0 = [n] \cup \{0\}$. As usual, $\oplus$ denotes the symmetric difference operator and $\delta(v)$ the neighborhood of a node v.

2. When spending does not help

2.1. The `Binary Unstable Matching` problem

We start with a lemma that relates the size of a matching to the number of edges blocking it.

Lemma 4. *Let* μ, S *be matchings of* $(G(V, E), >)$*, with* S *stable. Then* $|B_\mu^S| \geq |\mu| - |S|$.

Proof. For matchings μ', μ'' of G and $x \in V$, we define $vote(x, \mu', \mu'') = 1$ (resp. -1) if x prefers μ' to μ'' (resp. μ'' to μ'), and 0 otherwise. Let $vote(e, \mu', \mu'') := vote(m, \mu', \mu'') + vote(w, \mu', \mu'')$ for $e = mw \in E$. We moreover define $vote(\mu', \mu'') := \sum_{v \in V} vote(v, \mu', \mu'')$. It is known (see [10, Section 1.1]) that $vote(S, \mu) \geq 0$. Hence, by decomposing $vote(S, \mu)$ along the partition of V given by $(V(S), V(\mu) \setminus V(S), V \setminus (V(S) \cup V(\mu)))$, we have:

$$
\begin{aligned}
0 \quad \leq \quad & \underbrace{\sum_{v \in V(S)} vote(v, S, \mu)}_{= \sum_{e \in S} vote(e, S, \mu)} + \underbrace{\sum_{v \in V(\mu) \setminus V(S)} vote(v, S, \mu)}_{= -2|\mu \setminus S|} \\
& + \underbrace{\sum_{v \in V \setminus (V(S) \cup V(\mu))} vote(v, S, \mu)}_{=0} \\
\leq \quad & \sum_{e \in S} vote(e, S, \mu) - 2(|\mu| - |S|),
\end{aligned}
$$

where last inequality holds since $|\mu \setminus S| \geq |\mu| - |S|$.

Since $vote(e, S, \mu) \in \{0, \pm 2\}$ for all $e \in S$, there must be at least $|\mu| - |S|$ edges from S such that $vote(e, S, \mu) = 2$. All those are edges of S that block μ. □

Lemma 4 implies that, when $p(e) = 1$ for all $e \in E$, it is optimal to output a(ny) stable matching, as the weight gained by increasing the size of the matching is offset by the blocking edges that are introduced. As shown next, Lemma 4 also reveals that an optimal strategy for the `Binary Unstable Matching` problem is to not pay for any edge e with $p(e) = 1$ and to not include in the matching any edge e with $p(e) = 0$.

Lemma 5. *Let* $\mathcal{I}$ *be an instance of the* `Binary Unstable Matching` *problem and* $E_1 := \{e \in E : p(e) = 1\}$*. Then any stable matching of* $G(V, E_1)$ *is an optimal solution to* $\mathcal{I}$*.*

Proof. Let $E_0 := E \setminus E_1$ and, for a set $F \subseteq E_0$, let $G_F := G(V, E_1 \cup F)$. Let μ be a matching of G. Set $F(\mu) := \mu \cap E_0$ and let S' be a stable matching of $G_{F(\mu)}$. Then:

$$
\begin{aligned}
f(\mu) = p(\mu) - p(B_\mu) &= |\mu \cap E_1| - |B_\mu \cap E_1| \\
&= |\mu| - |F(\mu)| - |B_\mu \cap E(G_{F(\mu)})| \\
&\leq \cancel{|\mu|} - |F(\mu)| + |S'| - \cancel{|\mu|} \\
&\leq |S' \cap E_1| \\
&= p(S'),
\end{aligned}
\tag{1}
$$

where the chain of (in)equalities from the second row onwards follow respectively: by definition of $F(\mu)$ and $F(\mu) \cap B_\mu = \emptyset$; by applying Lemma 4 to graph $G_{F(\mu)}$ and matchings μ, S'; since $S' \cap E_0 \subseteq F(\mu)$; and by definition of p. Observing that no edge of $F(\mu) \cup E_1$ blocks S', we deduce:

$$p(S') = p(S') - p(B_{S'} \cap E_0) = p(S') - p(B_{S'}) = f(S').$$

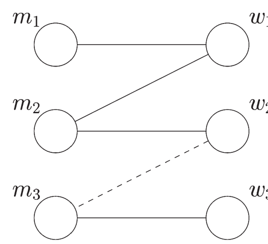

Fig. 1. Let $p(e) = 1$ for all edges e except the dashed edge f for which $p(f) = \beta \in (0, 1)$. w_1 and m_2 prefer each other to their other neighbors. Suppose w_2 prefers m_2 to m_3, and m_3 prefers w_2 to w_3. Let $\mu^* = \{m_1w_1, m_2w_2, m_3w_3\}$. Then $B_{\mu^*} = \{m_2w_1\}$ and $f(\mu^*) = p(\mu^*) - p(B_{\mu^*}) = 3 - 1 = 2$. μ^* is the unique optimal solution, since every other matching μ satisfies $f(\mu) < 2$. When $\beta \in \{0, 1\}$, μ^* is still optimal (e.g., by continuity), but Lemma 4 and Lemma 5 imply it is not the unique optimal solution.

The previous chain of equalities, together with (1), implies that there is an optimal solution to $\mathcal{I}$ in the form of a stable matching of G_F, for some $F \subseteq E_0$. Let now $S_\emptyset$ be a stable matching of $G(V, E_1) = G_\emptyset$ and S_F be a stable matching of G_F, $F \subseteq E_0$. Note that, for $S \in \{S_\emptyset, S_F\}$, we have $|V_S| = 2f(S)$, where $V_S := \{v \in V : E_1 \cap \delta(v) \cap S \neq \emptyset\}$.

Hence, in order to conclude the proof, it suffices to show that $|V_{S_F}| \leq |V_{S_\emptyset}|$. Suppose by contradiction this is not the case, and let $G' := (V, (E_1 \cap S_F) \oplus (E_1 \cap S_\emptyset))$. There must exist a connected component of G' with strictly more nodes from V_{S_F} than from $V_{S_\emptyset}$. Since each node from V has degree at most 2 in G', this connected component is a path $P = v_0, v_1, \ldots, v_t$, t odd, with v_0, v_t matched by $E_1 \cap S_F$ and exposed by $E_1 \cap S_\emptyset$, while the edges of P alternate between $E_1 \cap S_F$ and $E_1 \cap S_\emptyset$. Since $E_1 \cap S_\emptyset = S_\emptyset$, v_0 and v_t are unmatched in $S_\emptyset$. Hence, v_0 and v_t prefer S_F to $S_\emptyset$. If v_1 also prefers S_F to $S_\emptyset$, then v_0v_1 blocks $S_\emptyset$, contradicting the fact that $S_\emptyset$ is not blocked by any edge of E_1. Hence v_1 prefers $S_\emptyset$ to S_F. Since, by construction, S_F is also not blocked by any edge from E_1, we can iterate and deduce that v_i prefers S_F to $S_\emptyset$ for i even and $S_\emptyset$ to S_F for i odd. Thus, v_t prefers $S_\emptyset$ to S_F, a contradiction. □

Finally, we observe that Lemma 5 is not true even if $p(e) = 1$ for all edges e, except one edge f with $p(f) = \beta \in (0, 1)$. Indeed, Fig. 1 shows an instance of the β-`Binary Unstable Matching` problem for which the (unique) optimal solution is blocked by an edge e with $p(e) = 1$. This gives an intuition as to why the β-`Binary Unstable Matching` is substantially harder than the `Binary Unstable Matching` problem.

2.2. The `Preference-Concordant Unstable Matching` problem

Let $(G, >)$ be a preference system with $V(G) = \{m_1, w_1, m_2, w_2\}$ and $E(G) = \{m_iw_j\}_{i,j=1,2}$. Let $p(m_1w_1) = 6$, $p(m_1w_2) = p(m_2w_1) = 5$, $p(m_2w_2) = 1$. Under the unique $>$ that makes p preference-concordant, the optimal matching is $\{m_1w_1, m_2w_2\}$. If instead $w_2 >_{m_1} w_1$ and $w_1 >_{m_2} w_2$, the optimal matching is $\{m_1w_2, m_2w_1\}$. Hence, preference-concordancy cannot be assumed without loss of generality.

Algorithm 1 for the `Preference-Concordant Unstable Matching` problem iteratively finds a stable matching in the subgraph with edges of maximum weight and removes matched nodes from the graph. We show next that the union of all matchings constructed by the algorithm gives an optimal solution.

Lemma 6. *For every instance $\mathcal{I}$, Algorithm 1 outputs an optimal* `Preference-Concordant Unstable Matching` *on $\mathcal{I}$.*

Algorithm 1

Require: An instance $\mathcal{I}$ of the `Preference-Concordant Unstable Matching` problem: a preference system $(G(V, E), >)$, and a preference-concordant function $p : E \to \mathbb{Z}_{\geq 0}$.
Ensure: An optimal solution to $\mathcal{I}$.

```
Let S = ∅.
while G has still an edge do
    Let p* be the maximum weight of an edge of G and G' the subgraph of G
    that includes all and only the edges e with p(e) = p*.
    Let S̄ be a stable matching of G'.
    Set S = S ∪ S̄.
    Set V = V \ V(S̄).
end while
Output S.
```

Proof.

Claim 7. *S is a stable matching of $(G, >)$.*

Proof of claim. S is clearly a matching of G. Assume that S is not stable and let mw be an edge blocking it, i.e., $w >_m S(m)$ and $m >_w S(w)$. Let i be the iteration of Algorithm 1, in which the first among m and w is matched, i.e., m, w are both unmatched at the beginning of iteration i. Then, without loss of generality $mS(m) \in \overline{S}$ where $\overline{S}$ is a stable matching in the graph G' constructed at iteration i. As $mS(m)$ is an edge of maximum weight at iteration i and m, w are both unmatched at the beginning of iteration i, it must be that $p(mS(m)) \geq p(mw)$. Also, $w >_m S(m)$ yields through preference-concordancy that $p(mw) \geq p(mS(m))$. Then $p(mw) = p(mS(m))$, thus mw appears in graph G' of iteration i and blocks $\overline{S}$ as well, a contradiction. □

Fix a matching μ of G. We will show that $f(\mu) \leq f(S)$. Consider the graph $G_\mu(V, E_\mu) := G(V, S \oplus \mu)$ and let $\mathcal{C}$ be the set of its non-singleton connected components. For $U \subseteq V$, let $\Delta(U) := p(\mu(U)) - p(B^S_\mu(U)) - p(S(U))$, where for $F \subseteq E$, we let $F(U) := \{uv \in F : u, v \in U\}$. We argue that $\Delta(U) \leq 0$ for all $U \in \mathcal{C}$, where, to simplify matters, we denote by U both a member of $\mathcal{C}$ and its node set. Then, using Claim 7 we have $B_S = \emptyset$ and conclude:

$$\begin{aligned} f(\mu) - f(S) &= p(\mu) - p(B_\mu) - p(S) \\ &\leq p(\mu) - p(B^S_\mu) - p(S) \\ &= \textstyle\sum_{U \in \mathcal{C}} (p(\mu(U)) - p(B^S_\mu(U)) - p(S(U))) \\ &= \textstyle\sum_{U \in \mathcal{C}} \Delta(U) \\ &\leq 0. \end{aligned}$$

Suppose by contradiction that $\Delta(U) > 0$ for some $U \in \mathcal{C}$, and take a minimum counterexample, i.e., an instance $\mathcal{I}$ and matchings S, μ that violate the statement, so that the underlying graph G has a minimum number of nodes. Define $E_{\max} := \{e \in E_\mu(U) : p(e) \geq p(f) \text{ for all } f \in E_\mu(U)\}$ and $G_{\max} := G(U, E_{\max})$. If $G_{\max}$ is a cycle, then it has an even number of edges and $\Delta(U) = p(\mu(U)) - p(B^S_\mu(U)) - p(S(U)) \leq p(\mu(U)) - p(S(U)) = 0$, a contradiction. Otherwise, since $G_{\max}$ contains at least one edge and every node in U has degree at most 2 in G_μ (and hence in $G_{\max}$), we deduce that $G_{\max}$ is a path $P = v_0, v_1, \ldots, v_t$.

Claim 8. *$P \cap B^S_\mu \neq \emptyset$.*

Proof of claim. Suppose $v_0v_1 \in S$, as the other case is handled similarly. First observe that v_iv_{i+1} belongs to S for i even and to μ for i odd. As v_0 is an endpoint of P, then either v_0 is exposed in μ, or $p(v_0S(v_0)) > p(v_0\mu(v_0))$. In the latter case, by definition of preference-concordancy, $S(v_0) >_{v_0} \mu(v_0)$. Hence, in both cases, v_0 prefers S to μ.

Suppose there is an index j such that both v_j and v_{j+1} prefer the same matching between S and μ, and let j be the smallest such index. If j is odd, since v_0 prefers S to μ, then v_j and v_{j+1} prefer μ to S. We argued above that $v_j v_{j+1} \in \mu$. Hence, $v_j v_{j+1}$ blocks S, a contradiction to Claim 7. If conversely j is even, then v_j and v_{j+1} prefer S to μ, and $v_j v_{j+1} \in S$. Hence, we have found the required edge.

Assume then that nodes of P alternatively prefer S to μ and μ to S. If t is odd (resp. even), then $v_{t-1}v_t \in S$ (resp. μ), while v_t prefers μ to S (resp. S to μ). In both cases, we contradict the fact that v_t is an endpoint of P. $\square$

Using Claim 8, let $mw \in P \cap B^S_\mu$. Let U' (and possibly U'') be the connected components of the subgraph of G_μ induced by $U \setminus \{m, w\}$. We have:

$$\begin{aligned} 0 &< \Delta(U) \\ &= \Delta(U') + \Delta(U'') - 2p(mw) + p(m\mu(m)) + p(\mu(w)w) \\ &\leq \Delta(U') + \Delta(U''), \end{aligned}$$

where the equality follows since $mw \in B^S_\mu$ and the last inequality because $mw \in E_{\max}$ (note that U'' or $e \in \{m\mu(m), \mu(w)w\}$ may not exist, in which case we set $\Delta(U'') = 0$ or $p(e) = 0$). In particular, we can assume w.l.o.g. that $\Delta(U') > 0$.

Let $V' = V \setminus \{m, w\}$ and consider the subinstance $\mathcal{I}'$ of $\mathcal{I}$ on G', the subgraph of G induced by V'. Let $S' := S \setminus \{mw\}$, $\mu' := \mu \setminus \{m\mu(m), \mu(w)w\}$. Clearly, μ' is a matching of G'. It is easy to verify that there is an execution of Algorithm 1 that, on input $\mathcal{I}'$, outputs S'. For $U \subseteq V'$, let now $\Delta'(\cdot)$ be defined as $\Delta(\cdot)$, but for instance $\mathcal{I}'$ and matchings S', μ' instead of instance $\mathcal{I}$ and matchings S, μ. Observe that U' is a connected component of $G'(V', S' \oplus \mu')$, and $\Delta'(U') = \Delta(U') > 0$, contradicting the choice of $\mathcal{I}$ as a minimum counterexample. Hence, $\Delta(U) \leq 0$. $\square$

3. When spending gets hard

It will be convenient to work with an equivalent version of the β-`Binary Unstable Matching` defined as follows:

Given: A preference system $(G(V, E), >)$ and a function $p : E \to \{\alpha, \beta\}$ where $\alpha > \beta > 0$.
Find: A matching μ of $(G(V, E), >)$ maximizing $f(\mu)$.

Theorem 9. *The β-`Binary Unstable Matching` problem is NP-hard for any $\alpha > \beta > 0$ such that $\alpha > 4\beta$, even if preference lists are of length at most 3.*

We show Theorem 9 through a reduction from the special case of SAT used for `MMSBP` [4]. Unlike `MMSBP`, whose input is unweighted, we need to assign weights α or β to edges so that the β-`Binary Unstable Matching` instance has an optimal matching that is perfect and blocked by no edge of weight α.

For a Boolean formula $\mathcal{B}$ in CNF and a truth assignment h, $t(h)$ is the number of clauses of $\mathcal{B}$ satisfied by h and $t(\mathcal{B})$ the maximum value of $t(h)$ over all truth assignments. Let `MAX (2,2)-E3-SAT` [3] denote the NP-hard problem of finding a truth assignment h such that $t(h) = t(\mathcal{B})$, where each clause in $\mathcal{B}$ has exactly 3 literals and each variable occurs exactly twice as a non-negated literal and twice as a negated literal.

Let $v_1, \ldots, v_{n-1}$ and $c_1, \ldots, c_k$ be the set of variables and clauses of an instance $\mathcal{B}$ of `MAX (2,2)-E3-SAT` respectively, i.e., each clause has exactly three literals, and literals v_i and $\overline{v}_i$ appears

Table 1
Preference lists and pricing of instance $\mathcal{I}$.

	α	β	α
x_{6i+1} :	y_{6i+1}	$c(x_{6i+1})$	y_{6i+2}
x_{6i+2} :	y_{6i+2}	$c(x_{6i+2})$	y_{6i+3}
x_{6i+3} :	y_{6i+4}	$c(x_{6i+3})$	y_{6i+3}
x_{6i+4} :	y_{6i+5}	$c(x_{6i+4})$	y_{6i+4}
x_{6i+5} :		y_{6i+5}	y_{6i+6}
x_{6i+6} :		y_{6i+1}	y_{6i+6}
	β	α	α
y_{6i+1} :	x_{6i+6}	x_{6i+1}	
y_{6i+2} :		x_{6i+1}	x_{6i+2}
y_{6i+3} :		x_{6i+3}	x_{6i+2}
y_{6i+4} :		x_{6i+4}	x_{6i+3}
y_{6i+5} :	x_{6i+5}	x_{6i+4}	
y_{6i+6} :		x_{6i+5}	x_{6i+6}
	β	β	α
p^1_j :	w^1_j	c^1_j	
p^2_j :	w^2_j	c^2_j	
p^3_j :	w^3_j	c^3_j	
u^1_j :	w^1_j		z_j
u^2_j :	w^2_j		z_j
u^3_j :	w^3_j		z_j
c^1_j :	p^1_j	$x(c^1_j)$	q_j
c^2_j :	p^2_j	$x(c^2_j)$	q_j
c^3_j :	p^3_j	$x(c^3_j)$	q_j
w^1_j :	u^1_j	p^1_j	
w^2_j :	u^2_j	p^2_j	
w^3_j :	u^3_j	p^3_j	
	α	α	α
z_j :	u^1_j	u^2_j	u^3_j
q_j :	c^1_j	c^2_j	c^3_j

exactly twice. We form an instance $\mathcal{I}$ of the β-`Binary Unstable Matching` problem from $\mathcal{B}$ by introducing 6 men and 6 women per variable in $\mathcal{B}$ plus 7 men and 7 women per clause, as follows. The set of men in $\mathcal{I}$ is $P \cup U \cup X \cup Q$:

$$P := \{p^s_j : j \in [k], s \in [3]\},\ U := \{u^s_j : j \in [k], s \in [3]\},$$
$$X := \{x_{6i+r} : i \in [n-1]_0, r \in [6]\}, \text{ and } Q := \{q_j : j \in [k]\}.$$

The set of women is $C \cup W \cup Y \cup Z$ (to be consistent with [4], in this section W represents a subset of women):

$$C := \{c^s_j : j \in [k], s \in [3]\},\ W := \{w^s_j : j \in [k], s \in [3]\},$$
$$Y := \{y_{6i+r} : i \in [n-1]_0, r \in [6]\}, \text{ and } Z := \{z_j : j \in [k]\}.$$

We also make use of the subsets $X_i := \{x_{6i+r} : r \in [6]\}$,

$$Y_i := \{y_{6i+r} : r \in [6]\},\ P_j := \{p^1_j, p^2_j, p^3_j\},$$
$$U_j := \{u^1_j, u^2_j, u^3_j\},\ C_j := \{c^1_j, c^2_j, c^3_j\},\ W_j := \{w^1_j, w^2_j, w^3_j\},$$

defined for the appropriate indices i, j. The preference lists and the weights of the edges are given in Table 1.

In the preference list of a man $x_{6i+r} \in X$, $i \in [n]$ and $r \in \{1, 2\}$, the symbol $c(x_{6i+r})$ denotes the woman $c^s_j \in C$ such that the r-th occurrence of variable v_i appears at position s of clause c_j. Similarly, $c(x_{6i+r})$ for $r \in \{3, 4\}$ denotes the woman $c^s_j \in C$ such that the $(r-2)$-th occurrence of $\overline{v}_i$ appears at position s of c_j. Accordingly, $x(c^s_j)$ denotes the unique man in the preference list of c^s_j. Let us start with an observation that follows immediately from $\alpha, \beta > 0$.

Observation 10. *Every optimal solution for $\mathcal{I}$ is a maximal matching.*

Two important features regarding the weights of the edges in our construction are summarized next.

Observation 11. *If an edge mw blocks a matching μ and neither m nor w is single in μ then $p(mw) = \beta$, as $mw \in X \times C$ or $mw \in U \times W$ or $mw = x_{6i+5}y_{6i+5}$ or $mw = x_{6i+6}y_{6i+1}$ for some $i \in [n]$.*

Observation 12. *For $i \in [n]$, the only perfect matchings of $X_i \times Y_i$ are $T_i := \{x_{6i+\ell}y_{6i+\ell}\}_{\ell=1,\dots,6}$ and*

$F_i := \{x_{6i+\ell}y_{6i+\ell+1}\}_{\ell=1,\dots,5} \cup \{x_{6i+6}y_{6i+1}\}$.

Observe that $\overline{\mu} := \bigcup_{i=1}^n T_i \cup \{p_j^1c_j^1, p_j^2c_j^2, p_j^3w_j^3, u_j^1w_j^1, u_j^2w_j^2, u_j^3z_j, q_jc_j^3 : j \in [k]\}$ is a perfect matching of $\mathcal{I}$ that contains $5n+2k$ edges of weight α and is blocked by no edge of weight α (by Observation 11). Combined with $\alpha > 4\beta$, those observations imply the following two results.

Lemma 13. *An optimal matching for $\mathcal{I}$ contains no edge from $X \times C$.*

Proof. Let μ be an optimal matching and suppose, by contradiction, that there exists an edge $x_{6i+\ell}c_j^a \in \mu \cap (X \times C)$ for appropriate indices i, ℓ, a, j. Consider the matching

$$\mu' := \mu \setminus ((X_i \times Y_i) \cup \{x_{6i+\ell}c_j^a\} \cup S^-) \cup T_i \cup S^+,$$

where, $S^+ = \{c_j^a q_j\}$ and $S^- = \{\mu(q_j)q_j\}$ if $c_j^a >_{q_j} \mu(q_j)$, and $S^+ = S^- = \emptyset$ otherwise. μ' is clearly a matching. We show $f(\mu') > f(\mu)$, obtaining the required contradiction.

We claim $p(\mu' \cap (X_i \times Y_i)) - p(\mu \cap (X_i \times Y_i)) \geq \alpha$. Observe that $\mu' \cap (X_i \times Y_i) = T_i$ and $p(T_i) = 5\alpha + \beta$. Note that $\mu' \cap (X_i \times Y_i)$ contains at most 4 edges of weight α, since $x_{6i+\ell}$ is not matched via an edge of weight α, and both x_{6i+5}, x_{6i+6} are adjacent via an edge of weight α to y_{6i+6} only. Since $x_{6i+\ell}c_j^a \in \mu'$, we have $p(\mu \cap (X_i \times Y_i)) \leq 4\alpha + \beta$. The claim follows.

We now claim that $p(\mu' \setminus (X_i \times Y_i)) - p(\mu \setminus (X_i \times Y_i)) \geq -\beta$. If $S^+ = \emptyset$, no edge of weight α is removed from $\mu \setminus (X_i \times Y_i)$. Else, $\mu(q_j)q_j$ is removed and $c_j^a q_j$ is added, and both have weight α. Hence, the contribution of edges of weight α to $p(\mu' \setminus (X_i \times Y_i))$ and to $p(\mu \setminus (X_i \times Y_i))$ is the same. The only edge of weight β that is removed when moving from μ to μ' is $x_{6i+\ell}c_j^a$, and the claim follows.

We next argue that $p(B_{\mu'}) - p(B_\mu) \leq 3\beta$. Let e be an edge blocking μ', but not blocking μ. e must then be incident to a node that has changed partner between μ and μ'. If e is incident to $X_i \cup Y_i$, then $e \in \{x_{6i+3}c(x_{6i+3}), x_{6i+4}c(x_{6i+4}), x_{6i+6}y_{6i+1}\}$. We prove that e must be incident to $X_i \cup Y_i$, thus showing the claim. Suppose by contradiction it is not, then $e \in \{p_j^a c_j^a, q_j c_j^a\} \cup \delta(q_j) \cup \delta(\mu(q_j))$. If $e = p_j^a c_j^a$, then e also blocks μ, for p_j^a has not changed partner. By construction, $e \neq q_j c_j^a$. If e is incident to q_j but not to c_j^a, then it again also blocks μ. Last, note that e cannot be incident to $\mu(q_j)$, since q_j is the least favorite partner of $\mu(q_j)$.

Summing up, we obtain $f(\mu') - f(\mu) = p(\mu') - p(\mu) - p(B_{\mu'}) + p(B_\mu) \geq \alpha - \beta - 3\beta > 0$, as required. □

Lemma 14. *An optimal matching for $\mathcal{I}$ contains exactly $5n + 2k$ edges of weight α.*

Proof. Observe that a matching can contain up to 5 edges of weight α from $X_i \times Y_i, i \in [n]$, since each of x_{6i+5} and x_{6i+6} has exactly one incident edge of weight α, whose other endpoint is in both cases y_{6i+6}; this provides up to $5n$ edges of weight α from in $X \times Y$. Any other edge of weight α is incident to a node from $\{z_j, q_j\}_{j\in[k]}$, yielding up to k such edges from $Q \times C$ and k more from $U \times Z$ ($\overline{\mu}$ defined above attains all these maxima). Consider an optimal matching μ that contains less than $5n + 2k$ edges of weight α.

If μ contains less than 5 edges of weight α from $X_i \times Y_i$ for some $i \in [n]$, it may not contain edges of weight β from $X_i \times C$ (Lemma 13). Define the matching $\mu' := \mu \setminus (X_i \times Y_i) \cup T_i$. Then $p(\mu \cap (X_i \times Y_i)) \leq 4\alpha + 2\beta$, while $p(\mu' \cap T_i) = 5\alpha + \beta$. As in the proof of Lemma 13, at most 3 edges of weight β block μ' but not μ. This leads to the contradiction $f(\mu') - f(\mu) \geq (5\alpha + \beta - 3\beta) - (4\alpha + 2\beta) = \alpha - 4\beta > 0$.

If μ contains less than k edges of weight α from $Q \times C$, there exists $j \in [k]$ such that q_j is single in μ. As μ is maximal by Observation 10, all $c_j^r, r \in [3]$, are matched in μ. Since μ contains no edge in $X \times C$ (Lemma 13), $\mu(c_j^r) = p_j^r$ for each $r \in [3]$. The matching $\mu' := \mu \setminus \{p_j^3c_j^3\} \cup \{q_jc_j^3\}$ leaves p_j^3 single, introducing at most three edges blocking μ' but not μ, i.e., $p_j^3w_j^3$, $p_j^3c_j^3$ and $x(c_j^3)c_j^3$, all of weight β. Then, $f(\mu') - f(\mu) \geq \alpha - 3\beta - \beta > 0$, a contradiction.

If μ contains less than k edges of weight α from $U \times Z$, z_j is single in μ for some $j \in [k]$. As μ is maximal by Observation 10, $\mu(u_j^r) = w_j^r$ for each $r \in [3]$. The matching $\mu' := \mu \setminus \{u_j^3w_j^3\} \cup \{u_j^3z_j\}$ leaves w_j^3 single, thus introducing two edges blocking μ' but not μ, i.e., $u_j^3w_j^3$ and $p_j^3w_j^3$, both of weight β; again $f(\mu') - f(\mu) > 0$. □

Lemma 15. *There is an optimal matching in $\mathcal{I}$ that is perfect and induces a perfect matching in $X_i \times Y_i$ $\forall i \in [n]$.*

Lemma 15 is proved similarly to Lemma 13, hence we omit the details. We can now present the reduction.

Proof of Theorem 9. Let $f(\mathcal{I})$ denote the maximum value of $f(\mu)$ taken over all matchings μ of $\mathcal{I}$. We claim that $f(\mathcal{I}) = 5n\alpha + 3k\beta + 2k\alpha + t(\mathcal{B})\beta$ where we recall that $t(\mathcal{B})$ is the maximum value of $t(h)$, taken over all truth assignments h of the instance $\mathcal{B}$ of `MAX (2,2)-E3-SAT`. Let h be a truth assignment of $\mathcal{B}$ such that $t(h) = t(\mathcal{B})$. We create a perfect matching μ in $\mathcal{I}$ as in the reduction of `(3,3)-MMSBP` [4]. For each variable $v_i \in V$, if v_i is true under h, add the edges in T_i to μ, otherwise add the edges in F_i to μ. In both cases, it is easy to check that there is exactly one edge from $X_i \times Y_i$ blocking μ, respectively $x_{6i+6}y_{6i+1}$ or $x_{6i+5}y_{6i+5}$. Hence, $p(\mu \cap (X_i \times Y_i)) - p(B_\mu \cap (X_i \times Y_i)) = 5\alpha + \beta - \beta = 5\alpha$.

Now let $j \in [k]$. If c_j contains a literal that is true under h, let $s \in \{1, 2, 3\}$ denote the position of c_j in which this literal occurs, otherwise set $s = 1$. Add the edges $p_j^tc_j^t, u_j^tw_j^t$ $(1 \leq t \neq s \leq 3)$, $q_jc_j^s$, $p_j^sw_j^s$ and u_j^s, z_j to μ. The total weight of these edges is $2\alpha + 5\beta$. Moreover, $u_j^sw_j^s$ is a blocking edge and $p(u_j^sw_j^s) = \beta$. Now if c_j is not satisfied under h, then man $x(c_j^1)$ is assigned to his last-choice partner, by construction of μ. Hence $x(c_j^1)c_j^1$ is also a blocking edge and $p(x(c_j^1)c_j^1) = \beta$. Since μ is perfect, using Observation 11 one can check that, together with the $n + k$ blocking edges in $X \times Y$ and $U \times W$ identified already, these are all the blocking edges of μ in $\mathcal{I}$. Hence $f(\mathcal{I}) \geq f(\mu) = n(5\alpha) + k(2\alpha + 5\beta - \beta) - (k - t(\mathcal{B}))\beta = 5n\alpha + 2k\alpha + 3k\beta + t(\mathcal{B})\beta$. Conversely, let μ be an optimal matching for $\mathcal{I}$. We can assume by Lemma 15 that μ is perfect and that it induces a perfect matching of $X_i \times Y_i$, for every $i \in [n]$. By Observation 12, this matching is either F_i or T_i. Also, as there is no single agent in μ, by Observation 11 the edges blocking μ are in $X \times C$ or in $U \times W$ or of the form $x_{6i+5}y_{6i+5}$ and $x_{6i+6}y_{6i+1}$ and are all of weight β.

We form a truth assignment h in $\mathcal{B}$ through μ as follows. For each $i \in [n]$, if $\mu \cap (X_i \times Y_i) = T_i$, we set v_i to be true under h; otherwise, i.e., if $\mu \cap (X_i \times Y_i) = F_i$, we set v_i to be false. It also holds that $|B_\mu| = n + k + (k - t(h))$ (n edges of $X \times Y$, k edges of

$U \times W$ and $k - t(h)$ of $X \times C$), hence $f(\mu) \le 5n\alpha + n\beta + 2k\alpha + 5k\beta - (n + k + (k - t(\mathcal{B})))\beta = 5n\alpha + 2k\alpha + 3k\beta + t(\mathcal{B})\beta$. Given also our earlier inequality, it follows that $f(\mu) = 5n\alpha + 2k\alpha + 3k\beta + t(\mathcal{B})\beta$. Hence, an optimal solution to $\mathcal{I}$ is found if and only if `MAX (2,2)-E3-SAT` is solved to optimality, and the thesis follows. □

The hardness of approximation for `MAX (2,2)-E3-SAT` [3], combined with the above, leads to the following, whose proof follows that of [4, Theorem 7].

Corollary 16. *For every pair of constants α, β with $\alpha > 4\beta > 0$, the β-*`Binary Unstable Matching` *problem is* `APX-Hard`*.*

Given Corollary 16, it is natural to look for an algorithm whose approximation ratio depends on α and β. A relevant idea is to examine the proximity of a maximum weight stable matching (`MWSM`) to an optimal solution.

Lemma 17. *If μ' is an optimal matching for an instance of the β-*`Binary Unstable Matching` *problem where the image of p is $\{\beta, \alpha\}$, and μ is an* `MWSM`*, then $\frac{f(\mu')}{f(\mu)} \le 2 \cdot \frac{\alpha}{\beta} - 1$.*

Proof. Let $|\mu| = k$. Every edge from μ is of weight at least β hence $f(\mu) \ge k\beta$. It is well-known that every stable matching is maximal and every maximal matching is of cardinality at least half that of a maximum matching. Thus, μ' satisfies $|\mu'| \le 2k$ as such a matching cannot have more edges than a maximum matching. Then, Lemma 4 yields that at least $|\mu'| - k$ edges block μ'. Observe that $f(\mu')$ is maximized if every edge that belongs to μ' is of weight α and every edge that blocks μ' is of weight β. This implies that

$$\begin{aligned} f(\mu') &\le |\mu'|\alpha - (|\mu'| - k)\beta = |\mu'|(\alpha - \beta) + k\beta \\ &\le 2k(\alpha - \beta) + k\beta = 2k\alpha - k\beta. \end{aligned}$$

It follows that $\frac{f(\mu')}{f(\mu)} \le \frac{2k\alpha - k\beta.}{k\beta} = 2 \cdot \frac{\alpha}{\beta} - 1$. □

A small example shows that this bound is tight.

Example 1. Let $M = \{m_1, m_2\}$ and $W = \{w_1, w_2\}$, where m_1 prefers w_2 to w_1, w_2 prefers m_1 to m_2, m_2 wishes to be matched only to w_2 and w_1 only to m_1. Let also m_1w_1 and m_2w_2 be of weight α and m_1w_2 be of weight β. The only stable matching is $\mu = \{m_1w_2\}$ and has value $f(\mu) = \beta$. The optimal β-`Binary Unstable Matching` is $\mu' = \{m_1w_1, m_2w_2\}$ and has value $f(\mu') = 2\alpha - \beta$ as m_1w_2 is a blocking edge. Then, $\frac{f(\mu')}{f(\mu)} = 2 \cdot \frac{\alpha}{\beta} - 1$.

4. The `Red-Blue Unstable Matching` problem

Recall that the goal of the `Red-Blue Unstable Matching` problem is to find a matching μ that is stable in the subgraph induced only by the "blue edges" $E \setminus F$ and maximizes the sum of the weight of edges in μ minus the sum of the weight of "red edges" F blocking μ. We use classical concepts and facts related to rotations and the associated poset, referring the reader to [8] for a more extensive presentation.

Consider an instance $\mathcal{I}$ of the `Red-Blue Unstable Matching` problem and let $G' := G(V, E \setminus F)$, R be the set of rotations of $(G', >)$, and μ_0 its man-optimal stable matching. A rotation ρ is an ordered set of edges $(m_0w_0, \ldots, m_{r-1}w_{r-1})$ that, if *exposed* (i.e., identified) at some stable matching μ and then *eliminated* by shifting clockwise the men in ρ, yields another stable matching denoted as $\mu\Delta\rho$; i.e., $\mu\Delta\rho := \mu \setminus \{m_iw_i, i \in [r-1]_0\} \cup \{m_iw_{i+1}, i \in [r-1]_0\}$, indices taken modulo r. For each stable matching μ, there is exactly one set of rotations $R' = \{\rho_1, \ldots, \rho_k\} \subseteq R$ whose successive elimination from the man-optimal matching yields μ, i.e., $\mu = \mu_0\Delta R' := ((\mu_0\Delta\rho_1)\ldots\Delta\rho_k)$. The rotation poset $(R, \succeq)$ is defined as follows: for $\rho, \rho' \in R$, we have $\rho \succeq \rho'$ if and only if, for any R' such that $\mu_0\Delta R'$ is a stable matching and $\rho' \in R'$, we have $\rho \in R'$. The following holds.

Theorem 18. *[11, Theorem 4.1] There is a bijection between the set of stable matchings of $(G', >)$ and the closed sets of $(R, \succeq)$, mapping each closed set R' of $(R, \succeq)$ to the stable matching $\mu_0\Delta R'$.*

We now extend weights to rotations [12]. For $\rho = (m_0w_0, \ldots, m_{r-1}w_{r-1})$, let $p(\rho) := \sum_{i=0}^{r-1} p(m_iw_{i+1}) - p(m_iw_i)$. Then, if $\mu = \mu_0\Delta R'$, we have $p(\mu) = p(\mu_0) + p(R')$. Let $R^+ := \{\rho \in R : p(\rho) \ge 0\}$ and $R^- := R \setminus R^+$. We recall that when eliminating a rotation ρ exposed at μ, every man (resp. woman) either obtains in $\mu\Delta\rho$ a less preferred (resp. more preferred) partner, or (s)he does not change partner. We can then associate certain rotations to each red edge.

Observation 19. *Let $m \in M$, $w \in W$ and $e = mw \in F$.*

- *There is at most one $\rho^+(e) = (m_0w_0, \ldots, m_{r-1}w_{r-1}) \in R$ such that $w = w_i$ and $m_{i-1} >_w m >_w m_i$ for some $i \in [r-1]_0$.*
- *There is no $\rho = (m_0w_0, \ldots, m_{r-1}w_{r-1}) \in R$ such that $w = w_i$ and $m_i >_w m >_w m_{i-1}$ for some $i \in [r-1]_0$.*
- *There is at most one $\rho^-(e) = (m_0w_0, \ldots, m_{r-1}w_{r-1}) \in R$ such that $m = m_i$ and $w_i >_m w >_m w_{i+1}$ for some $i \in [r-1]_0$.*
- *There is no $\rho = (m_0w_0, \ldots, m_{r-1}w_{r-1}) \in R$ such that $m = m_i$ and $w_{i+1} >_m w >_m w_i$ for some $i \in [r-1]_0$.*

The next labeling is an adaptation to our case of one frequently used for *popular matching* problems (see, e.g., [10]). We attach to $mw \in F$ with $m \in M$ and $w \in W$ a label given by an ordered pair (ℓ_m, ℓ_w), defined as follows:

- If $\mu_0(m) >_m w$, then set $\ell_m = -$, else set $\ell_m = +$;
- If $\mu_0(w) >_w m$, then set $\ell_w = -$, else set $\ell_w = +$.

We next write $e = (+, +)$ (or e is $(+, +)$) to mean that the label associated to e as above is $(+, +)$ (and similarly for other labels).

We now have all the necessary ingredients to sketch the idea of our algorithm. μ_0 is blocked by all and only $(+, +)$ edges. Recall that, when we eliminate an exposed rotation, no man improves his partner and no woman worsen her partner. Hence, as we iteratively eliminate exposed rotations starting from μ_0, some $(+, +)$ edges will be "deactivated", i.e., they become $(+, -)$ and do not block the current matching, while certain $(-, +)$ edges will be "activated", i.e., they become $(+, +)$ and block the current matching. Thus, we need to find a stable matching μ in $(G', >)$ – or equivalently, by Theorem 18, a closed set of rotations from $(R, \succeq)$ – so that $p(\mu)$ minus the sum of $p(e)$ for all "activated" edges e plus the sum of $p(e)$ for all "deactivated" edges e is maximized. We compute such a matching by solving a minimum cut problem in a digraph (similarly to the classical algorithm for computing a stable matching of maximum weight [12], the main difference being the presence of edges between two nodes corresponding to rotations in our digraph).

We now employ the ideas above to partition F.

F_0 (resp. F_1) is the set of edges $e \in F$ with $e = (+, +)$ and such that $\rho^+(e)$ does not exist (resp. exists).
F_2 (resp. F_3) is the set of edges $e \in F$ with $e = (-, +)$ and such that both $\rho^+(e)$ and $\rho^-(e)$ exist (resp. ρ^+ does not exist, while ρ^- exists);

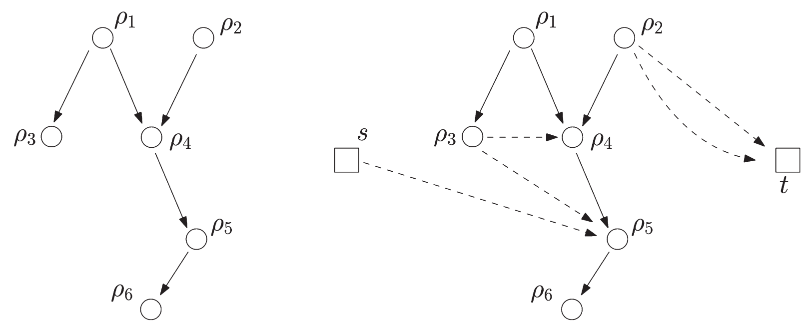

Fig. 2. On the left: A (rotation) poset. On the right: An example of the multidigraph $D(U, K)$ (edges from K_{R^+}, K_{R^-} and those that are implied by transitivity of $\succeq$ are omitted). Full edges have infinite capacity, while each of the dashed edges has capacity $p(e)$ for some $e \in E$. Here $F_1 = \{e_1, e_2\}$ with $\rho^+(e_1) = \rho^+(e_2) = \rho_2$; $F_2 = \{e_3, e_4\}$ with $\rho^+(e_3) = \rho^+(e_4) = \rho_3$, $\rho^-(e_3) = \rho_4$, $\rho^-(e_4) = \rho_5$; $F_3 = \{e_5\}$ with $\rho^-(e_5) = \rho_5$. The cut $\{s, \rho_3, \rho_5, \rho_6\}$ of capacity 1 corresponds to the closed set $\{\rho_1, \rho_2, \rho_4\}$.

F_4 is the set $F \setminus \cup_{i=0}^{3} F_i$.

With respect to the intuitive description given above, $e \in F_0$ is $(+,+)$ and never "deactivated", $e \in F_1$ is $(+,+)$ and "deactivated" if and only if we eliminate $\rho^+(e)$, $e \in F_2$ (resp. $e \in F_3$) is $(-,+)$ and "activated" if and only if we eliminate $\rho^-(e)$ but not $\rho^+(e)$ (resp., we eliminate $\rho^-(e)$), while no edge of F_4 is ever "activated". This is formalized in the next lemma.

Lemma 20. *Let μ be a stable matching of $(G', >)$ and let $\mu = \mu_0 \Delta R'$ for some $R' \subseteq R$. Then:*

0. *All edges of F_0 block μ;*
1. *An edge $e \in F_1$ blocks μ if and only if $\rho^+(e) \notin R'$;*
2. *An edge $e \in F_2$ blocks μ if and only if $\{\rho^+(e), \rho^-(e)\} \cap R' = \{\rho^-(e)\}$;*
3. *An edge $e \in F_3$ blocks μ if and only if $\rho^-(e) \in R'$;*
4. *No edge of F_4 blocks μ.*

Proof. We show only 0., as the remainder follows similarly. Let $e = mw \in F_0$. By definition of F_0, $w >_m \mu_0(m)$ and $m >_w \mu_0(w)$. By definition of μ_0, $\mu_0(m) \geq_m \mu(m)$. On the other hand, since $\rho^+(e)$ does not exist, we have $m >_w \mu(w)$, hence e blocks μ. □

We next construct a capacitated multidigraph $D(T, K)$ with $T = R \cup \{s, t\}$ and K the disjoint union of sets:

$K_0 := \{(\rho, \rho')\}_{\rho, \rho' \in R \text{ with } \rho \succ \rho'}$, each with capacity $+\infty$;
$K_1 := \{(\rho^+(e), t)\}_{e \in F_1}$, each with capacity $p(e)$;
$K_2 := \{(\rho^+(e), \rho^-(e))\}_{e \in F_2}$, each with capacity $p(e)$;
$K_3 := \{(s, \rho^-(e)\}_{e \in F_3}$, each with capacity $p(e)$;
$K_{R^+} := \{(\rho, t)\}_{\rho \in R^+}$, each with capacity $p(\rho)$;
$K_{R^-} := \{(s, \rho)\}_{\rho \in R^-}$, each with capacity $-p(\rho)$.

Let w be the vector of capacities defined above. Note that $w \geq 0$ and within each of K_1, K_2, K_3, multiple copies of an arc may be present, see Fig. 2.

Algorithm 2 clearly runs in polynomial time, and next lemma shows that it solves the `Red-Blue Unstable Matching` problem. As it is customary, for $S \subseteq T$, we define $\delta^+(S)$ to be the set of arcs outgoing from nodes in S in D.

Lemma 21. *Let S be an $s-t$ cut of $D(T, K)$ of finite capacity w.r.t. w. Then $R \setminus S$ is a closed set of $(R, \succeq)$. Moreover, $w(\delta^+(S)) = p(\mu_0) - p(F_0) + p(R^+) - f(\mu)$ where $\mu := \mu_0 \Delta (R \setminus S)$. Conversely, for every closed set R' of $(R, \succeq)$, $S := \{s\} \cup (R \setminus R')$ is an $s-t$ cut.*

Proof. Let S be an $s-t$ cut with finite capacity, and let $\rho \succ \rho' \in R$. As the arc (ρ, ρ') has infinite capacity, we have that $\rho' \in (R \setminus S)$ implies $\rho \in (R \setminus S)$, showing that $R \setminus S$ is closed. Hence, $\mu = \mu_0 \Delta (R \setminus S)$ is a stable matching. Conversely, since the only arcs with infinite capacity are (ρ, ρ') with $\rho \succ \rho' \in R$, we deduce that, if R' is closed, then $\{s\} \cup (R \setminus R')$ is an $s-t$ cut of finite capacity. By Theorem 18, this argument defines a bijection between $s-t$ cuts S of finite capacity in $D(T, K)$ and stable matchings of $(G, >)$. For an $s-t$ cut S of finite capacity, we have:

$$\begin{aligned} w(\delta^+(S)) = {} & p(\underbrace{e \in F_1 : \rho^+(e) \in S}_{(1)}) \\ & + p(\underbrace{e \in F_2 : \rho^+(e) \in S, \rho^-(e) \in R \setminus S}_{(2)}) \\ & + p(\underbrace{e \in F_3 : \rho^-(e) \in R \setminus S}_{(3)}) \\ & + p(R^+ \cap S) - p(R^- \setminus S). \end{aligned} \tag{2}$$

Let μ_S be the matching associated to S via the bijection defined above. By Lemma 20, an edge $e \in F$ blocks μ_S if and only if it belongs to $F_0 \cup (1) \cup (2) \cup (3)$. We deduce:

$$\begin{aligned} f(\mu) &= p(\mu) - p(B_\mu) \\ &= p(\mu_0) + p(R \setminus S) - w(\delta^+(S)) \\ &\quad + p(R^+ \cap S) - p(R^- \cap S) - p(F_0) \\ &= p(\mu_0) - w(\delta^+(S)) - p(F_0) \\ &\quad \underbrace{+ p(R^+ \cap S) + p(R \setminus S) - p(R^- \setminus S)}_{(*)}, \end{aligned}$$

where the second equality follows by the discussion above, (2) and $p(\mu) = p(\mu_0) + p(R \setminus S)$. Rearranging and observing $(*) = p(R^+ \cap S) + p(R^+ \setminus S) = p(R^+)$ concludes the proof. □

Algorithm 2

Require: An instance $\mathcal{I}$ of the `Red-Blue Unstable Matching` problem: a preference system $(G(V, E), >)$, a function $p : E \to \mathbb{Z}_{\geq 0}$ and a set $F \subseteq E$.
Ensure: An optimal solution to $\mathcal{I}$.

1: Compute the man-optimal stable matching μ_0 from $(G(V, E \setminus F), >)$, the set R of rotations, and its corresponding poset $(R, \succeq)$, see e.g. [8].
2: Label the edges of F, build the partition $(F_0, \ldots, F_4)$ of F and the digraph $D(T, K)$ with edge capacities w computed as above.
3: Find a minimum $s-t$ cut S^* in $D(T, K)$.
4: **output** $\mu = \mu_0 \Delta (R \setminus S^*)$.

It is not hard to see how to modify the construction above to solve the `Red-Blue Unstable Matching` when the nonnegativity assumption on p is dropped: we need to replace each of K_1, K_2, K_3 with two sets, depending on the sign of $p(e)$, and modify Algorithm 2 and Lemma 21 appropriately.

Acknowledgements

We thank the reviewer for their insightful comments and suggestions. Yuri Faenza acknowledges support from NSF award 2046146.

References

[1] Atila Abdulkadiroğlu, Tayfun Sönmez, House allocation with existing tenants, J. Econ. Theory 88 (2) (1999) 233–260.
[2] Georgios Askalidis, Nicole Immorlica, Augustine Kwanashie, David Manlove, Emmanouil Pountourakis, Socially stable matchings in the hospitals/residents problem, in: Algorithms and Data Structures (WADS2013), 2013, pp. 85–96.
[3] Piotr Berman, Marek Karpinski, Alex Scott, Approximation Hardness of Short Symmetric Instances of MAX-3SAT, Electronic Colloquium on Computational Complexity (ECCC), 2003.
[4] Péter Biró, David Manlove, Shubham Mittal, Size versus stability in the marriage problem, Theor. Comput. Sci. 411 (16) (2010) 1828–1841.
[5] Lars Ehlers, Thayer Morrill, (Il)legal assignments in school choice, Rev. Econ. Stud. (2019).
[6] Yuri Faenza, Xuan Zhang, Legal assignments and fast EADAM with consent via classical theory of stable matchings, arXiv preprint, arXiv:1809.08506, 2018.
[7] David Gale, Lloyd S. Shapley, College admissions and the stability of marriage, Am. Math. Mon. 69 (1) (1962) 9–15.
[8] Dan Gusfield, Robert W. Irving, The Stable Marriage Problem: Structure and Algorithms, MIT Press, 1989.
[9] Koki Hamada, Kazuo Iwama, Shuichi Miyazaki, An improved approximation lower bound for finding almost stable maximum matchings, Inf. Process. Lett. 109 (18) (2009) 1036–1040.
[10] Chien-Chung Huang, Telikepalli Kavitha, Popular matchings in the stable marriage problem, Inf. Comput. 222 (2013) 180–194.
[11] Robert W. Irving, Paul Leather, The complexity of counting stable marriages, SIAM J. Comput. 15 (3) (1986) 655–667.
[12] Robert Irving, Paul Leather, Dan Gusfield, An efficient algorithm for the "optimal" stable marriage, J. ACM 34 (1987) 532–543.