Stable fractional matchings ☆

Ioannis Caragiannis [a], Aris Filos-Ratsikas [b], Panagiotis Kanellopoulos [c],
Rohit Vaish [d],*

[a] *Department of Computer Science, Aarhus University, Åbogade 34, 8200 Aarhus N, Denmark*
[b] *Department of Computer Science, University of Liverpool, Liverpool, L69 3BX, United Kingdom*
[c] *School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, United Kingdom*
[d] *School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, 400005, India*

ARTICLE INFO

ABSTRACT

We study a generalization of the classical stable matching problem that allows for *cardinal* preferences (as opposed to ordinal) and *fractional* matchings (as opposed to integral). In this cardinal setting, stable fractional matchings can have much larger social welfare than stable integral ones. Our goal is to understand the computational complexity of finding an *optimal* (i.e., welfare-maximizing) stable fractional matching. We consider both *exact* and *approximate* stability notions, and provide simple approximation algorithms with weak welfare guarantees. Our main result is that, somewhat surprisingly, achieving better approximations is computationally hard. To the best of our knowledge, these are the first computational complexity results for stable fractional matchings in the cardinal model. En route to these results, we provide a number of structural observations that could be of independent interest.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The stable matching problem is one of the most extensively studied problems at the interface of economics and computer science [1–5]. The input to the problem consists of the preference lists of two sets of agents, commonly referred to as the *men* and the *women*. The goal is to find a *stable* matching, i.e., a matching in which no pair of man and woman prefer each other over their assigned partners.

While the problem was originally motivated by college admissions [1], its applicability has subsequently expanded to various other domains such as medical residency [6,7] and school choice [8]. In addition, the insights gained from the study of the stable matching problem, together with the development of computational tools and techniques in artificial intelligence, have shaped the design of modern two-sided matching platforms such as organ exchanges [9] and ridesharing platforms [10].

The standard formulation of the stable matching problem involves two important assumptions, namely that the matching is *integral* (i.e., two agents are either completely matched or completely unmatched) and that the agents have *ordinal* preferences (typically in the form of rank-ordered lists). Although these assumptions suffice in a number of applications,

---

☆ A preliminary version of this article has appeared in the proceedings of the 2019 ACM Conference on Economics and Computation (pp. 21-39).
* Corresponding author.

E-mail addresses: caragian@ceid.upatras.gr (I. Caragiannis), arisfilosratsikas@gmail.com (A. Filos-Ratsikas), panagiotis.kanellopoulos@essex.ac.uk (P. Kanellopoulos), rohit.vaish@tifr.res.in (R. Vaish).
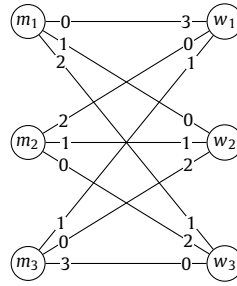
Fig. 1. An instance with cardinal preferences.

including those mentioned above, there are natural examples where they could be inadequate. For instance, consider a *time-sharing* scenario [11] wherein a set of employees are matched with a set of supervisors. Assuming that each individual can spend one unit of time at work, an integral matching prescribes that every employee should work full time with a single supervisor. On the other hand, fractional matchings allow the employees to divide their time in working with multiple supervisors, making them a more natural modeling choice in such situations. Fractional matchings are also useful in the context of *randomization*, as they can be used to model lotteries over integral matchings [12–14].

In a similar vein, *ordinal* preferences, despite their simplicity and ease of elicitation, can often be quite restrictive. Indeed, in many real-world matching applications, the outcomes experienced by the participants are inherently *cardinal* in nature (e.g., wages in labor markets or quality of transplants in kidney exchange [15]). In such settings, it is decidedly more natural to model the *intensity* of preferences, as has been noted both in theory [16,17] as well as in lab experiments [18].

Motivated by these applications, we consider a generalization of the stable matching model that allows for *fractional* matchings (as opposed to integral) and *cardinal* preferences (as opposed to ordinal). More concretely, we consider a setting in which the preferences are specified in terms of numerical utilities or *valuations* (for example, in the matching instance in Fig. 1, m_1 values w_1 at 0, and w_1 values m_1 at 3). A fractional matching is simply a convex combination of integral matchings, and an agent's utility under a fractional matching is the appropriately weighted sum of its utilities under the constituent integral matchings. A fractional matching μ is *stable* if no pair of man and woman simultaneously derive greater utility in being integrally matched to each other than they do under μ [19,17]. Thus, for instance, in the employee-supervisor example mentioned above, stability ensures that no employee-supervisor pair will abandon their time-sharing arrangements and instead prefer to work with each other full time.

The aforementioned generalization has clear merit in terms of *social welfare*: Stable solutions in the generalized model can have larger welfare than those in the standard model, as the following example illustrates.

Example 1. Consider the instance in Fig. 1 with three men m_1, m_2, m_3 and three women w_1, w_2, w_3. Among the six possible integral matchings, only two are stable, namely $\mu_1 := \{(m_1, w_3), (m_2, w_2), (m_3, w_1)\}$ and $\mu_2 := \{(m_1, w_1), (m_2, w_3), (m_3, w_2)\}$; indeed, μ_1 and μ_2 are the men-proposing and women-proposing Gale-Shapley matchings, respectively [1]. The social welfare (i.e., the sum of utilities of all agents) of these matchings is $\mathcal{W}(\mu_1) = \mathcal{W}(\mu_2) = 7$.

Define $\mu_3 := \{(m_1, w_1), (m_2, w_2), (m_3, w_3)\}$, and notice that $\mathcal{W}(\mu_3) = 8$. Now consider a *fractional* matching $\mu := \frac{1}{2}\mu_2 + \frac{1}{2}\mu_3$. The social welfare of μ is $\mathcal{W}(\mu) = \frac{1}{2}\mathcal{W}(\mu_2) + \frac{1}{2}\mathcal{W}(\mu_3) = 15/2 > 7 = \mathcal{W}(\mu_1) = \mathcal{W}(\mu_2)$, which means that μ has a higher social welfare than any stable integral matching. Importantly, μ is a *stable* fractional matching. Indeed, in μ, the utilities of $m_1, m_2, m_3, w_1, w_2,$ and w_3 are 0, 1/2, 3/2, 3, 3/2, and 1 respectively. Thus, for every man-woman pair, at least one of the two agents meets its utility threshold for that pair, implying that μ is stable.

Overall, the instance in Fig. 1 admits a stable fractional matching with strictly greater welfare than any stable integral matching.[1] □

Starting with the seminal work of Gale and Shapley [1], an extensive literature has emerged over the years on algorithms for computing stable solutions, including ones that optimize a variety of objectives pertaining to fairness and economic efficiency [20,21,4,22,23]. Many of these algorithms, however, are tailored to compute stable *integral* matchings. As Example 1 demonstrates, such algorithms could, in general, return highly suboptimal outcomes in our setting. Therefore, it becomes pertinent to understand the computational complexity of finding an "optimal" stable matching in the generalized model. Our work studies this question from the lens of the fundamental objective of social welfare, and asks the following natural question:

Can an (approximately) optimal stable fractional matching be efficiently computed?

---

[1] Observe that the gain in social welfare of μ was achieved by including an *unstable* integral matching μ_3 in its support. One could consider alternative notions such as *ex-post stability* [11] wherein the support consists only of stable integral matchings. We discuss this and various other stability notions in Appendix A, and find that insisting on a purely stable support results in significant welfare loss (Remark 3). In fact, there exists instances where the support of an optimal stable fractional matchings consists only of unstable integral matchings (Proposition 4).

1.1. Our results and roadmap

We formalize the above question by defining the optimization problem Optimal Stable Fractional Matching. To motivate this problem, we strengthen the observation in Example 1 to show that the social welfare gap between the best stable fractional and the best stable integral matchings can be *arbitrarily* large. We show that the favorable welfare properties of stable fractional matchings come at the cost of limiting the algorithmic tools at our disposal. Specifically, we show that the set of stable fractional matchings can be *non-convex* (Proposition 3), and that, in the worst case, stable fractional matchings can have a *large support* (Theorem 3), thus prohibiting the use of support enumeration algorithms. Nevertheless, we present simple algorithms for Optimal Stable Fractional Matching with approximation ratio of $1 + \sigma_{\max}/\sigma_{\min}$, where $\sigma_{\max}$ and $\sigma_{\min}$ represent the maximum and minimum positive valuation in the input instance, respectively (Theorem 5). For the variant Optimal ε-Stable Fractional Matching, where the stability constraints are relaxed by a multiplicative factor of $(1 - \varepsilon)$, an embarrassingly simple algorithm computes $1/\varepsilon$-approximate solutions (Theorem 6). We then proceed to our main results (Theorems 7 and 8), which show that these approximation guarantees are—somewhat surprisingly—almost the best achievable via polynomial-time algorithms (unless P = NP). To the best of our knowledge, these are the first computational complexity results for stable fractional matchings in the cardinal preferences model.

The rest of the paper is organized as follows. We present related work in the matching literature in Section 1.2. We continue in Section 2 with preliminary definitions and warm up with exponential-time algorithms that solve Optimal Stable Fractional Matching using linear programming. Section 3 presents the structural properties of (nearly)-optimal solutions of Optimal Stable Fractional Matching. Our algorithms are presented in Section 4, and our inapproximability results are presented in Section 5. Some proofs and additional material appear in the appendix.

1.2. Related work

The stable matching problem has been extensively studied for *integral* matchings. The universal existence of a stable integral matching [1] has led to considerable research on stable solutions that, in addition, optimize various measures of fairness or efficiency. A relevant example is that of optimizing the *average rank* of matched partners, which, in our model, corresponds to finding an optimal stable integral matching when the cardinal utilities are completely specified by the ordinal ranks (e.g., if an agent is ranked at position i, then it is valued at $n + 1 - i$). This problem is known to admit combinatorial polynomial-time algorithms [21,24]. Other examples of such problems include minimizing the difference between average ranks of matched partners of men and women [25], optimizing the rank of matched partner for the least well-off agent [20, 26,4], and maximizing (or minimizing) the cardinality of the matching [27,23,28].

A growing body of work in artificial intelligence and multiagent systems has studied variants of the stable matching problem motivated by practical considerations such as minimizing the amount of information exchange required in arriving at stable outcomes [29,30] or ensuring strategyproofness in many-to-one matchings with quota constraints [31,32]. Various other papers have used computational approaches such as SAT solving [33–35] and constraint programming [36–38] in developing practically efficient solutions to computationally hard variants of the problem, as well as in axiomatic study of matching procedures [39].

Stability has also been studied in the context of *fractional* matchings. Starting with the works of Vande Vate [40], Rothblum [41], and Roth et al. [11], there is now a well-developed literature on linear programming formulations of the stable matching problem [22,42,43]. This line of work has led to novel stability notions such as *strong stability* [11], *ex-post stability* [11], and *fractional stability* [40], which are discussed further in Appendix A.[2] A noteworthy difference with our work is that these notions have been studied with respect to purely *ordinal* preferences.

Turning to *cardinal* preferences, we note that the notions of exact and approximate stability studied by us (Definition 1) first appeared, to the best of our knowledge, in the work of Anshelevich et al. [19]. They focus on *qualitative*, rather than computational, questions such as analyzing the "price of anarchy" for integral stable matchings under various preference structures as well as its extensions to approximate stability.[3] Pini et al. [17] study the same notion with the goal of computing integral stable matchings that, in addition, satisfy economic efficiency (in particular, Pareto optimality and its variants). They also study strategic aspects which are an exciting avenue for future research even in our model.

An interesting special case of our problem is when agents have *symmetric* preferences, i.e., for every $(m, w) \in M \times W$, $U(m, w) = V(m, w)$. Deligkas et al. [47] study this model in the context of *integral* matchings, and show that computing a welfare-maximizing integral stable matching is NP-hard under symmetric valuations. In general, stable fractional matchings can have much higher welfare compared to integral ones (Example 1), and thus it is not clear a priori whether the result of Deligkas et al. for integral matchings has any implications for stable fractional matchings. Nevertheless, as we show in Lemma 1 in Section 2.2, their result implies NP-hardness of Optimal Stable Fractional Matching. In comparison, our results in Section 5 on the hardness of approximation are much stronger and also apply to approximate stability.

---

[2] The term "fractional stable matching" has been overloaded in the literature. For example, Teo and Sethuraman [22] use it to refer to a feasible solution of stable matching linear program, and Aharoni and Fleiner [44] and Biró and Fleiner [45] use it in the study of hypergraphic preference systems to refer to a slightly different solution concept. We refer the reader to [46] for a detailed overview of these notions.

[3] The price of anarchy in this context is the worst-case multiplicative welfare gap between an optimal matching and an integral stable matching.

Finally, we note that stable *fractional* matchings under cardinal preferences have been previously studied in economics literature [48–52]. Some of these works [49,50] consider much more general matching models such as matching with transfers, blocking coalitions of arbitrary size, non-linear utilities, etc. The focus in these papers is primarily on *existential* questions (such as the existence of competitive equilibria) or on examining the *logical* relationship among various notions of stability and economic efficiency. Some of these models are strict generalizations of ours, and therefore computational hardness results in our model readily extend to these more general settings. Whether one can obtain stronger inapproximability results for these models is an interesting avenue for future research.

2. Preliminaries

An instance of *Stable Matching problem with Cardinal preferences* (SMC) is given by the tuple $\langle M, W, U, V \rangle$, where $M :=$ $\{m_1, \ldots, m_n\}$ and $W := \{w_1, \ldots, w_n\}$ denote the set of n men and n women, respectively, and U and V are $n \times n$ matrices of non-negative rational numbers that specify the *valuations* of the agents. Specifically, $U(m, w)$ is the value derived by man m from his match with woman w, and $V(m, w)$ is the value derived by woman w from her match with man m. Many of our results will focus on two special classes of valuations, namely *binary* (where $U, V \in \{0, 1\}^{n \times n}$) and *ternary* valuations (where $U, V \in \{0, 1, \alpha\}^{n \times n}$ for some $\alpha > 1$).

We will often describe an SMC instance using its *graph representation*. An instance $\mathcal{I} = \langle M, W, U, V \rangle$ can be represented as a bipartite graph with vertex sets M and W, and an edge for every pair $(m, w) \in M \times W$ such that at least one of $U(m, w) > 0$ or $V(m, w) > 0$ holds. Each edge (m, w) in this graph has two valuations associated with it, namely $U(m, w)$ and $V(m, w)$.

A *fractional matching* $\mu : M \times W \to \mathbb{R}_{\geq 0}$ is an assignment of non-negative weights to all man-woman pairs such that $\sum_{w \in W} \mu(m, w) \leq 1$ for each $m \in M$ and $\sum_{m \in M} \mu(m, w) \leq 1$ for each $w \in W$. A fractional matching μ is said to be *complete* if $\sum_{w \in W} \mu(m, w) = 1$ for each man $m \in M$ and $\sum_{m \in M} \mu(m, w) = 1$ for each woman $w \in W$. An *integral* matching μ is a fractional matching with weights $\mu(m, w) \in \{0, 1\}$ for every pair (m, w). With slight abuse of notation, we sometimes view an integral matching μ as a set of pairs and write $(m, w) \in \mu$ in place of $\mu(m, w) = 1$. Also, unless stated otherwise, we will assume that any fractional/integral matching is complete.

It is well-known, and follows from the Birkhoff-von Neumann (BvN) theorem, that a (complete) fractional matching μ can be written as a convex combination of $k = \mathcal{O}(n^2)$ integral matchings $\mu^{(1)}, \mu^{(2)}, \ldots, \mu^{(k)}$ so that for every pair $(m, w) \in M \times W$, we have

$$\mu(m, w) = \sum_{j=1}^{k} \lambda_j \cdot \mu^{(j)}(m, w),$$

where $\lambda_j > 0$ for all $j \in \{1, \ldots, k\}$ and $\sum_{j=1}^{k} \lambda_j = 1$. The set of integral matchings $\{\mu^{(1)}, \ldots, \mu^{(k)}\}$ is called the *support* of the fractional matching μ. Note that the support need not be unique.

We proceed with the formal definitions of *stability* and *approximate stability*, which, in turn, depend on the *utility* derived by agents in a fractional matching. In particular, the utility derived by the man m in μ is given by $u_m(\mu) := \sum_{w \in W} U(m, w)\mu(m, w)$, and the utility derived by the woman w is given by $v_w(\mu) := \sum_{m \in M} V(m, w)\mu(m, w)$.

Definition 1 (*Stability [19,17]*). Given a fractional matching μ, a man-woman pair (m, w) is said to be a *blocking pair* if $u_m(\mu) < U(m, w)$ and $v_w(\mu) < V(m, w)$. A fractional matching μ is *stable* if there are no blocking pairs, i.e., for each $(m, w) \in M \times W$, either $u_m(\mu) \geq U(m, w)$ or $v_w(\mu) \geq V(m, w)$.

Thus, under a stable fractional matching, no pair of man and woman can simultaneously improve by breaking away from the fractional matching and instead being integrally matched with each other. This notion of deviation is also reasonable from the viewpoint of bounded rationality, as agents only form blocking coalitions of size two, and only deviate to an integral matching between the members of the coalition.

Definition 2 (*ε-Stability [19]*). Given any $\varepsilon \in [0, 1)$ and a fractional matching μ, a man-woman pair (m, w) is said to be ε-*blocking* if $u_m(\mu) < (1 - \varepsilon)U(m, w)$ and $v_w(\mu) < (1 - \varepsilon)V(m, w)$; otherwise, the pair is said to be ε-*stable*. A fractional matching μ is ε-*stable* if all pairs are ε-stable.

Thus, a 0.01-stable fractional matching is one in which, for every man-woman pair, at least one of the two agents already receives (at least) 99% of the utility that he or she would receive by being integrally matched with the other. Note that a stable fractional matching is also ε-stable for every $\varepsilon \geq 0$.

Notice that Definitions 1 and 2 entail that agents in a blocking pair prefer to switch to an *integral* match with each other. One could also consider an alternative formulation wherein the agents merely prefer to *increase* their mutual fractional engagement, possibly at the expense of weakening other less preferred matches. This is precisely the notion of *strong stability* (see Appendix A for the definition) which has been studied in the context of ordinal preferences [11,46]. However,

as we note in Remark 3 in Appendix A, a strongly stable fractional matching can be strictly suboptimal in terms of social welfare, which further justifies the consideration of the stability notion in Definition 1.

The next statement follows from the seminal result of Gale and Shapley [1].

Proposition 1. *Given any* SMC *instance* $\mathcal{I}$, *a stable fractional matching* μ *for* $\mathcal{I}$ *always exists and can be computed in polynomial time.*

Proposition 1 was originally proven in [1] in the standard stable matching model with ordinal preferences and integral matchings. It is easy to see that given any SMC instance $\mathcal{I}$, if an integral matching μ is stable for an ordinal instance derived from $\mathcal{I}$ (where the ordinal preferences of each agent are consistent with its valuations, breaking ties arbitrarily), then it is also stable for the original instance $\mathcal{I}$.

Next, we define *social welfare*, which is a measure of the efficiency of a fractional matching.

Definition 3 *(Social welfare).* Given an SMC instance $\langle M, W, U, V \rangle$ and a fractional matching μ, the *social welfare* of μ is defined as

$$\mathcal{W}(\mu) := \sum_{m \in M} u_m(\mu) + \sum_{w \in W} v_w(\mu) = \sum_{m \in M} \sum_{w \in W} (U(m, w) + V(m, w))\mu(m, w).$$

An *optimal* matching is one with the highest social welfare among all fractional matchings. It follows from the BvN decomposition that an optimal matching is, without loss of generality, integral. Similarly, an *optimal stable* fractional matching (respectively, *optimal ε-stable* fractional matching) is one with the highest social welfare among all stable (respectively, all ε-stable) fractional matchings. We will use Optimal Stable Fractional Matching and Optimal ε-Stable Fractional Matching to refer to the corresponding optimization problems.

We will also discuss stable fractional matchings that are *approximately optimal*.

Definition 4 *(ρ-efficiency).* For $\rho \in (0, 1]$, the term *ρ-efficient* will refer to a stable (respectively, ε-stable) fractional matching with welfare at least ρ times the welfare of the optimal stable (respectively, ε-stable) fractional matching. That is, μ is a ρ-efficient stable fractional matching if it is stable (respectively, ε-stable) and $\mathcal{W}(\mu) \geq \rho \mathcal{W}(\mu^*)$, where μ^* is an optimal stable (respectively, ε-stable) fractional matching.

Thus, an optimal stable (or ε-stable) fractional matching is 1-efficient.

2.1. Computing optimal stable fractional matchings

We will now discuss two exponential-time algorithms for Optimal Stable Fractional Matching. The first algorithm uses the following mixed integer linear program (OPT-Stab):

(OPT-Stab)

$$\text{maximize} \quad \sum_{m \in M} u_m + \sum_{w \in W} v_w$$

$$\text{subject to} \quad u_m \geq U(m, w)y(m, w) \qquad \forall m \in M, w \in W \tag{1}$$

$$v_w \geq V(m, w)(1 - y(m, w)) \qquad \forall m \in M, w \in W \tag{2}$$

$$u_m = \sum_{w \in W} U(m, w)\mu(m, w) \qquad \forall m \in M \tag{3}$$

$$v_w = \sum_{m \in M} V(m, w)\mu(m, w) \qquad \forall w \in W \tag{4}$$

$$\sum_{w \in W} \mu(m, w) \leq 1 \qquad \forall m \in M \tag{5}$$

$$\sum_{m \in M} \mu(m, w) \leq 1 \qquad \forall w \in W \tag{6}$$

$$\mu(m, w) \geq 0 \qquad \forall m \in M, w \in W \tag{7}$$

$$y(m, w) \in \{0, 1\} \qquad \forall m \in M, w \in W \tag{8}$$

The non-negative weights $\mu(m, w)$ of man-woman pairs as well as the utilities $u_m := u_m(\mu)$ and $v_w := v_w(\mu)$ of the agents (set in equalities (3) and (4)) are the fractional variables of (OPT-Stab). The binary variables $y(m, w)$ encode the stability requirements for pair (m, w) in constraints (1) and (2). Indeed, by setting $y(m, w)$ to 1 or 0, we can require

either $u_m(\mu) \geq U(m, w)$ or $v_w(\mu) \geq V(m, w)$. Constraints (5) and (6) ensure feasibility. By enumerating over all possible combinations of values for the binary variables $y(m, w)$ for $(m, w) \in M \times W$, we get 2^{n^2} different linear programs, and at least one of them must have the optimal stable fractional matching as its optimal solution.

Our second algorithm is slightly faster and solves at most $\mathcal{O}(n^n)$ linear programs. It exploits the following linear program (OPT-Thresh), which is defined using non-negative constants θ_m for $m \in M$ and θ_w for $w \in W$, which we call *utility thresholds*.

(OPT-Thresh)

$$\text{maximize} \quad \sum_{m \in M} u_m + \sum_{w \in W} v_w$$

$$\text{subject to} \quad u_m \geq \theta_m \qquad\qquad\qquad \forall\, m \in M \tag{9}$$

$$v_w \geq \theta_w \qquad\qquad\qquad \forall\, w \in W \tag{10}$$

$$u_m = \sum_{w \in W} U(m, w)\mu(m, w) \qquad\qquad \forall\, m \in M \tag{11}$$

$$v_w = \sum_{m \in M} V(m, w)\mu(m, w) \qquad\qquad \forall\, w \in W \tag{12}$$

$$\sum_{w \in W} \mu(m, w) \leq 1 \qquad\qquad\qquad \forall\, m \in M \tag{13}$$

$$\sum_{m \in M} \mu(m, w) \leq 1 \qquad\qquad\qquad \forall\, w \in W \tag{14}$$

$$\mu(m, w) \geq 0 \qquad\qquad\qquad \forall\, m \in M, w \in W \tag{15}$$

When all utility thresholds are set to zero, the solution of (OPT-Thresh) is an optimal (i.e., welfare-maximizing) fractional matching. Using (OPT-Thresh) to maximize social welfare under stability constraints is more challenging. We say that a set of utility thresholds is *stability-preserving* if for every pair of agents $m \in M$ and $w \in W$, either $\theta_m \geq U(m, w)$ or $\theta_w \geq V(m, w)$. Note that any fractional matching μ that is feasible for (OPT-Stab) is also feasible for (OPT-Thresh) for some stability-preserving set of utility thresholds (in particular, set the threshold of an agent equal to its utility under μ). Conversely, any fractional matching μ that is feasible for (OPT-Thresh) with some set of stability-preserving utility thresholds is also feasible for (OPT-Stab).

One could now adopt the following strategy to solve OPTIMAL STABLE FRACTIONAL MATCHING: First, enumerate all $\mathcal{O}(n^n)$ tuples of utility thresholds $(\theta_{m_1}, \ldots, \theta_{m_n})$ with $\theta_m \in \{U(m, w) : w \in W\}$ for every man $m \in M$. Next, for every choice of $(\theta_{m_1}, \ldots, \theta_{m_n})$, solve (OPT-Thresh) after appropriately setting $(\theta_{w_1}, \ldots, \theta_{w_n})$ where $\theta_w \in \{V(m, w) : m \in M\}$ for all $w \in W$, so that the set of utility thresholds is stability-preserving. Among these solutions, the fractional matching with highest social welfare will be the solution of OPTIMAL STABLE FRACTIONAL MATCHING. We note that (OPT-Thresh) resembles the integer programming formulations with cut-off variables in other works [53,54].

2.2. The case of symmetric valuations

As mentioned previously in Section 1.2, a result by Deligkas et al. [47] implies NP-hardness of the problem of computing an optimal stable *integral* matching on SMC instances. The construction of Deligkas et al. involves a restricted class of SMC instances wherein the agents have *symmetric* (i.e., $U = V$) and *ternary* valuations in $\{0, 1, \alpha\}$ with $\alpha \in (1, 2)$.

Proposition 2. *Computing an optimal stable integral matching is NP-hard even for* SMC *instances with symmetric ternary valuations.*

Our next result (Lemma 1) shows that for these SMC instances, the optimal stable *fractional* matching is, without loss of generality, *integral*. Consequently, the result in [47] implies NP-hardness of OPTIMAL STABLE FRACTIONAL MATCHING.

Lemma 1. *Let $\mathcal{I}$ be an* SMC *instance with symmetric and ternary valuations, and let μ^* be an optimal stable fractional matching for $\mathcal{I}$. Then, there exists a stable integral matching μ^s such that $\mathcal{W}(\mu^s) = \mathcal{W}(\mu^*)$.*

Proof. Consider an optimal stable fractional matching μ^* that is not integral (i.e., has support of size at least two), and let μ be any integral matching in any support of μ^*. We will show that μ is a stable integral matching. This would imply that all matchings in the support of μ^* are optimal stable integral matchings, as desired.

Assume, for contradiction, that μ is not stable, and let (m, w) be a blocking pair in μ. Then, because of symmetry, either $U(m, w) = V(m, w) = 1$ or $U(m, w) = V(m, w) = \alpha$.
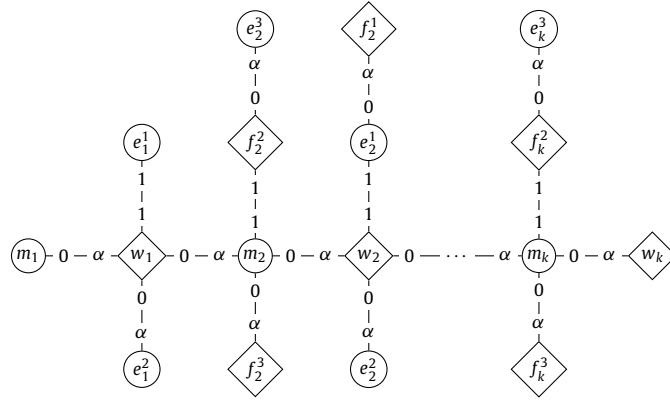
Fig. 2. The SMC instance used in the proof of Theorem 1. As a convention, in graph representations where the two sides of the bipartition do not appear as left and right, we use circles to represent men and diamonds to represent women.

Suppose that $U(m, w) = V(m, w) = 1$. Then, either both m and w are unmatched in μ, or any pair (m', w') in μ with either $m' = m$ or $w' = w$ satisfies $U(m', w') = V(m', w') = 0$. By replacing such pairs with (m, w) in μ (and, subsequently, in the support of μ^*), we get a stable fractional matching with a higher welfare than μ^*, contradicting its optimality.

Now suppose that $U(m, w) = V(m, w) = \alpha$. Then, either both m and w are unmatched in μ, or any pair (m', w') in μ with either $m' = m$ or $w' = w$ satisfies $U(m', w') = V(m', w') \in \{0, 1\}$. However, since μ is in the support of μ^* (i.e., with strictly positive weight) and all valuations are less than or equal to α, the utilities of both m and w in μ^* will be strictly smaller than α, contradicting the stability of μ^*.   □

From Lemma 1 and Proposition 2, we immediately have the following:

Corollary 1. OPTIMAL STABLE FRACTIONAL MATCHING *is NP-hard.*

3. Structural properties

In this section, we present several observations about the structure of optimal and nearly-optimal stable fractional matchings. We begin by considerably strengthening our observation in Example 1 regarding the welfare gap between stable fractional and stable integral matchings.

Theorem 1. *For every $\delta > 0$ and $\alpha \geq 2$, there exists an SMC instance with ternary valuations in $\{0, 1, \alpha\}$ and an optimal stable fractional matching μ^* such that any stable integral matching μ^s satisfies $\mathcal{W}(\mu^s) \leq \left(\alpha - \frac{1}{2} - \delta\right)^{-1} \mathcal{W}(\mu^*)$.*

Proof. Consider the SMC instance shown in Fig. 2, which, for some k (to be determined later), consists of man m_i and woman w_i for all $i \in \{1, \ldots, k\}$, men e_i^1 and e_i^2 for all $i \in \{1, \ldots, k-1\}$, woman f_i^1 for all $i \in \{2, \ldots, k-1\}$, man e_i^3 and women f_i^2 and f_i^3 for all $i \in \{2, \ldots, k\}$.

To specify the valuations, we will use the following notation: For any $a, b \geq 0$, we will say that a man-woman pair (m, w) is an "$a-b$" edge if $U(m, w) = a$ and $V(m, w) = b$. In the instance in Fig. 2, the pairs $\{(m_i, w_i)\}_{i=1}^{k}$, $\{(m_i, f_i^3)\}_{i=2}^{k}$, and $\{(e_i^1, f_i^1)\}_{i=2}^{k-1}$ are $0-\alpha$ edges, the pairs $\{(m_{i+1}, w_i)\}_{i=1}^{k-1}$, $\{(e_i^2, w_i)\}_{i=1}^{k-1}$, and $\{(e_i^3, f_i^2)\}_{i=2}^{k}$ are $\alpha-0$ edges, and the pairs $\{(e_i^1, w_i)\}_{i=1}^{k-1}$ and $\{(m_i, f_i^2)\}_{i=2}^{k}$ are $1-1$ edges. All other pairs are $0-0$ edges. We remark that a slightly modified version of the instance in Fig. 2 will be used again later in the proof of Theorem 7 (refer to the *accumulator gadget* in Fig. 7a).

Consider a stable integral matching μ^s. Since the pair (e_1^1, w_1) is a $1-1$ edge, the stability requirement dictates that either (m_1, w_1) or (e_1^1, w_1) is contained in μ^s. Either of these pairs contribute at most α to the social welfare (recall that $\alpha \geq 2$). Additionally, since μ^s is an integral matching, we get that $\mu^s(m_2, w_1) = 0$. This, in turn, forces the pair (m_2, f_2^2) to be contained in μ^s as well (or, otherwise, both m_2 and f_2^2 will have zero utility and violate stability). Continuing in this manner, we observe that the stability requirement for the pairs $\{(m_i, f_i^2)\}_{i=2}^{k}$ and $\{(e_i^1, w_i)\}_{i=2}^{k-1}$ will force these pairs to be contained in μ as well. Each of these pairs contributes 2 to the social welfare, and thus, $\mathcal{W}(\mu^s) \leq 4k - 6 + \alpha$.

Now define a stable *fractional* matching μ as follows: Set $\mu(m_i, w_i) = 1/\alpha$ for all $i \in \{1, \ldots, k\}$, $\mu(e_i^2, w_i) = 1 - 2/\alpha$, $\mu(m_{i+1}, w_i) = 1/\alpha$, $\mu(e_i^1, w_i) = 0$ for all $i \in \{1, \ldots, k-1\}$, $\mu(m_i, f_i^2) = 0$, $\mu(m_i, f_i^3) = 1 - 2/\alpha$, $\mu(e_i^3, f_i^2) = 1$ for all $i \in \{2, \ldots, k\}$, and $\mu(e_i^1, f_i^1) = 1$, for all $i \in \{2, \ldots, k-1\}$. (A similar matching will be used later in the proof of Lemma 5.) It is easy to verify that the social welfare of μ is $\mathcal{W}(\mu) = 4k(\alpha - 1/2) - 5\alpha + 3$. Furthermore, for each $i \in \{1, \ldots, k\}$, both man m_i and woman w_i have utility 1 in μ (i.e., $u_{m_i}(\mu) = v_{w_i}(\mu) = 1$). Every $1-1$ edge includes one of these agents, and these are the only edges that need to be checked for stability (any other pair has at least one agent with a valuation of zero and,
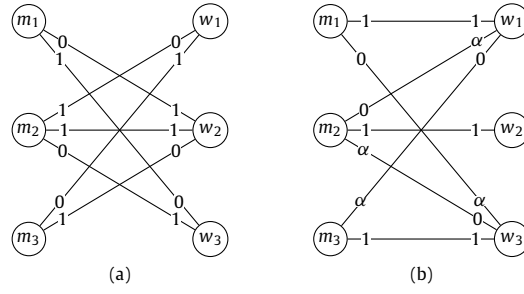
Fig. 3. The SMC instances used in the proofs of Propositions 3 and 4.

hence, the stability constraint for those pairs is trivially satisfied). Therefore, μ is stable. The theorem follows by setting k to be sufficiently large; specifically, for any given $\delta > 0$, choosing $k \geq \frac{1}{4} \cdot \left(6 - \alpha + \frac{\alpha^2 - 23\alpha/2 + 6}{\delta}\right)$ gives the desired bound. $\qquad\square$

We emphasize that Theorem 1 is a *positive* result as it establishes that stable fractional matchings can have much higher welfare than their integral counterparts, and highlights the importance of OPTIMAL STABLE FRACTIONAL MATCHING.

Our next observation (Proposition 3) shows that the set of stable fractional matchings can be *non-convex* even for binary valuations. Interestingly, this does not prove to be a barrier in efficiently solving OPTIMAL STABLE FRACTIONAL MATCHING in this setting (see Theorem 4 in Section 4).

Proposition 3. *There exists an SMC instance with binary valuations for which the set of stable fractional matchings is non-convex.*

Proof. Consider the instance $\mathcal{I} = \langle M, W, U, V \rangle$ with three men m_1, m_2, m_3 and three women w_1, w_2, w_3, whose graph representation and agent valuations are shown in Fig. 3a. Consider the integral matchings $\mu^{(1)} := \{(m_1, w_3), (m_2, w_1), (m_3, w_2)\}$ and $\mu^{(2)} := \{(m_1, w_2), (m_2, w_3), (m_3, w_1)\}$. It is easy to verify that both $\mu^{(1)}$ and $\mu^{(2)}$ are stable for $\mathcal{I}$. However, the fractional matching $\mu := 0.5\mu^{(1)} + 0.5\mu^{(2)}$ is not stable since (m_2, w_2) is a blocking pair; indeed, $0.5 = u_{m_2}(\mu) < U(m_2, w_2) = 1$ and $0.5 = v_{w_2}(\mu) < V(m_2, w_2) = 1$. $\qquad\square$

Remark 1. A follow-up work to the conference version of our paper has shown that non-convexity also holds for strict preferences [55].

Remark 2. It is worth comparing the non-convexity results in Proposition 3 and Remark 1 with similar results for other stability notions. In particular, Aziz and Klaus [46, Theorem 1] have shown that the set of strongly stable matchings (see Appendix A for the definition) can be non-convex. In Appendix B, we revisit the counterexample used in their proof, which is stated in terms of ordinal preferences, and show that there exists a realization of cardinal preferences consistent with their instance such that the relevant convex combination of (strongly) stable matchings is still stable. Thus, the non-convexity result of Aziz and Klaus [46] for strong stability does not imply the same for the set of stable matchings.

The structure of stable fractional matchings becomes even more interesting (and, as we will see in Section 5, also computationally unwieldy) when we move to *ternary* valuations. It turns out that the support of a stable fractional matching can comprise entirely of *unstable* integral matchings (Proposition 4), and its size can grow *linearly* with the input (Theorem 3). These observations pose major limitations on the set of algorithmic tools at our disposal.

Proposition 4. *There exists an SMC instance with ternary valuations and a stable fractional matching μ such that every integral matching in any support of μ is unstable.*

Proof. Consider the SMC instance $\mathcal{I} = \langle M, W, U, V \rangle$ with three men and three women shown in Fig. 3b. The parameter $\alpha \geq 3$ is a constant. There are six different perfect integral matchings:

- Matching $\mu^{(1)}$, which consists of pairs (m_1, w_1), (m_2, w_2), and (m_3, w_3) and has a social welfare of 6. It is easy to verify that this is the unique stable integral matching. Also, any subset of $\mu^{(1)}$ is not stable as the pair that is missing from $\mu^{(1)}$ will be blocking.
- Matching $\mu^{(2)}$, which consists of pairs (m_1, w_2), (m_2, w_3), and (m_3, w_1) and has a social welfare of 2α. The matching is not stable since the pair (m_1, w_1) is blocking.
- Matching $\mu^{(3)}$, which consists of pairs (m_1, w_3), (m_2, w_1), and (m_3, w_2) and has a social welfare of 2α. It is not stable since (m_2, w_2) is blocking.
- Matching $\mu^{(4)}$, which consists of pairs (m_1, w_1), (m_2, w_3), and (m_3, w_2) and has a social welfare of $\alpha + 2$. It is not stable since (m_3, w_3) is blocking.
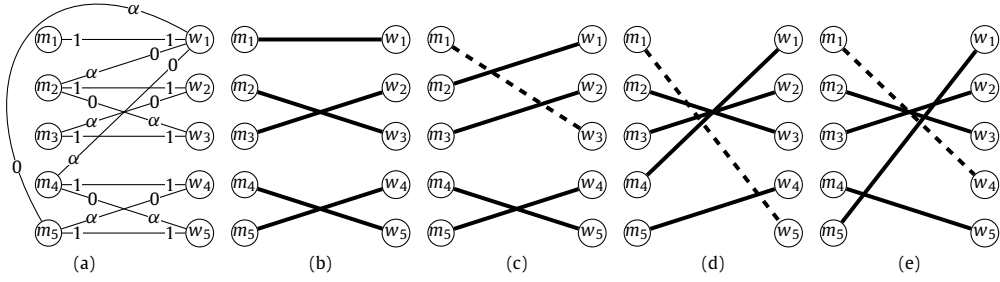
Fig. 4. Subfigure (a) illustrates the graph representation of the SMC instance $\mathcal{I}_n$ described in the proof of Theorem 3 for $n = 5$. Subfigure (b) shows the matching μ^{opt}. Subfigures (c), (d), and (e) show the matchings $\mu^{(1)}$, $\mu^{(2)}$, and $\mu^{(3)}$, respectively. Dashed lines indicate zero-valuation pairs that do not appear in the graph representation.

- Matching $\mu^{(5)}$, which consists of pairs (m_1, w_3), (m_2, w_2), and (m_3, w_1) and has a social welfare of $2\alpha + 2$. It is not stable since (m_1, w_1) is blocking.
- Matching $\mu^{(6)}$, which consists of pairs (m_1, w_2), (m_2, w_1), and (m_3, w_3) and has a social welfare of $\alpha + 2$. It is not stable since the pair (m_2, w_2) is blocking.

Consider the matching $\mu := \frac{1}{\alpha(\alpha-1)} \cdot \mu^{(2)} + \frac{1}{\alpha} \cdot \mu^{(3)} + \frac{\alpha-2}{\alpha-1} \cdot \mu^{(5)}$. It is easy to verify that μ is stable. Indeed, the utilities of the agents in μ are given by $u_{m_1}(\mu) = 0$, $u_{m_2}(\mu) = 1$, $u_{m_3}(\mu) = \alpha - 1$, $v_{w_1}(\mu) = 1$, $v_{w_2}(\mu) = \frac{\alpha-2}{\alpha-1}$ and $v_{w_3}(\mu) = \alpha - \frac{1}{\alpha-1}$. Notice that only the pairs (m_1, w_1), (m_2, w_2), and (m_3, w_3) need to be checked for stability. For each of these pairs, at least one of the agents has a utility of at least 1 in μ, implying that μ is stable. Finally, notice that $\mu(m_1, w_1) = 0$, which means that the unique stable integral matching $\mu^{(1)}$ cannot occur in any support of μ. □

We remark that with some extra work, one can show that the matching μ in the proof of Proposition 4 is the unique optimal stable fractional matching.

One might wonder whether the support of an optimal stable fractional matching always consists of an optimal integral matching. This turns out to not be the case, as illustrated by the following instance: Consider two men m_1, m_2 and two women w_1, w_2 with valuations given by $U = \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}$ and $V = \begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix}$. The unique optimal integral matching is $\{(m_1, w_1), (m_2, w_2)\}$ while the unique optimal stable fractional matching is $\{(m_1, w_2), (m_2, w_1)\}$.

As mentioned previously in Section 2, a (stable) fractional matching is the convex combination of at most n^2 integral ones. Theorem 2 provides a stronger bound on the support size of an *optimal* stable fractional matching.

Theorem 2. *Given any* SMC *instance* $\mathcal{I}$, *there exists an optimal stable fractional matching for* $\mathcal{I}$ *with at most $4n$ integral matchings in its support.*

Proof. Let μ^* be an optimal stable fractional matching for $\mathcal{I}$. Recall from Section 2.1 that μ^* solves the program (OPT-Thresh) for some set of stability-preserving utility thresholds. Observe that (OPT-Thresh) has n^2 free variables (we ignore here the $2n$ variables u_m for $m \in M$ and v_w for $w \in W$, which depend on the remaining ones according to constraints (11) and (12)). Without loss of generality, μ^* is an optimal *extreme point* solution of (OPT-Thresh). That is, when (OPT-Thresh) is instantiated for μ^*, n^2 linearly independent inequality constraints become tight. Among them, at most $4n$ can correspond to the sets of constraints (9), (10), (13), and (14). The remaining ones must correspond to the set of constraints (15), implying that at least $n^2 - 4n$ free variables will be equal to zero. Thus, μ^* can assign positive weights to at most $4n$ man-woman pairs and, consequently, can have at most $4n$ integral matchings in its support. □

Next we show that the bound in Theorem 2 is tight up to a constant factor.

Theorem 3. *For every $\rho \in (0, 1]$, there exists a family of* SMC *instances with ternary valuations for which any support of a ρ-efficient stable fractional matching consists of $\Omega(\rho n)$ integral matchings.*

Proof. Consider a family of SMC instances $\mathcal{I}_n = \langle M, W, U, V \rangle$ with $M = \{m_1, \ldots, m_n\}$ and $W = \{w_1, \ldots, w_n\}$, where n is odd. Let α be such that $\alpha > \max\left\{n+2, \frac{2n}{\rho(n-1)}\right\}$. The (ternary) valuations of the agents are defined as follows: For each $i \in \{1, 2, \ldots, n\}$, $U(m_i, w_i) = V(m_i, w_i) = 1$. For each $i \in \{1, 2, \ldots, \frac{n-1}{2}\}$, $U(m_{2i}, w_1) = U(m_{2i+1}, w_{2i}) = V(m_{2i}, w_{2i+1}) = \alpha$ and $V(m_{2i}, w_1) = V(m_{2i+1}, w_{2i}) = U(m_{2i}, w_{2i+1}) = 0$. Finally, $U(m_n, w_1) = 0$ and $V(m_n, w_1) = \alpha$. For all remaining pairs $(m, w) \in M \times W$, $U(m, w) = V(m, w) = 0$. Fig. 4a illustrates the SMC instance $\mathcal{I}_5$.

Define $\mu^{\text{opt}} := \{(m_1, w_1)\} \cup \left\{(m_{2i}, w_{2i+1}), (m_{2i+1}, w_{2i}) : i \in \{1, 2, \ldots, \frac{n-1}{2}\}\right\}$ (see Fig. 4b). We also define a number of other integral matchings obtained by modifying μ^{opt}, as follows: For all $i \in \{1, 2, \ldots, \frac{n-1}{2}\}$, the matching $\mu^{(i)}$ (see

Figs. [4]c and [4]d) is the integral matching which is obtained from μ^{opt} by replacing $\{(m_1, w_1), (m_{2i}, w_{2i+1})\}$ with $\{(m_1, w_{2i+1}), (m_{2i}, w_1)\}$, i.e.,

$$\mu^{(i)} := \Big\{ (m_1, w_{2i+1}) \cup (m_{2i}, w_1) \cup (m_{2i+1}, w_{2i}) \cup \{(m_{2\ell}, w_{2\ell+1}) \cup (m_{2\ell+1}, w_{2\ell})\}_{\ell \in \{1,2,\ldots,\frac{n-1}{2}\} \setminus \{i\}} \Big\}.$$

Also, the matching $\mu^{\left(\frac{n+1}{2}\right)}$ (see Fig. [4]e) is the integral matching obtained from μ^{opt} by replacing $\{(m_1, w_1), (m_n, w_{n-1})\}$ with $\{(m_1, w_{n-1}), (m_n, w_1)\}$, i.e.,

$$\mu^{\left(\frac{n+1}{2}\right)} := \Big\{ (m_1, w_{n-1}) \cup (m_{n-1}, w_n) \cup (m_n, w_1) \cup \{(m_{2\ell}, w_{2\ell+1}) \cup (m_{2\ell+1}, w_{2\ell})\}_{\ell \in \{1,2,\ldots,\frac{n-3}{2}\}} \Big\}.$$

Now, consider the fractional matching $\mu := \sum_{i=1}^{\frac{n+1}{2}} \frac{1}{\alpha}\mu^{(i)} + \left(1 - \frac{n+1}{2\alpha}\right)\mu^{\mathrm{opt}}$. Since $\alpha > n+2 > \frac{n+1}{2}$, μ is well-defined and has the matchings μ^{opt} and $\mu^{(i)}$ for all $i \in \{1, \ldots, \frac{n+1}{2}\}$ in its support. Notice that $\mathcal{W}(\mu^{\mathrm{opt}}) = (n-1)\alpha + 2$ and $\mathcal{W}(\mu^{(i)}) = (n-1)\alpha$ for all $i \in \{1, 2, \ldots, \frac{n+1}{2}\}$. Thus, $\mathcal{W}(\mu) > (n-1)\alpha$.

It can be verified that μ is stable. Indeed, we only need to check the blocking condition for the pairs (m_i, w_i) with $i \in \{1, 2, \ldots, n\}$. We have that $v_{w_1}(\mu) \geq 1$ (since $\mu^{\left(\frac{n+1}{2}\right)}$ has weight $\frac{1}{\alpha}$ in μ and $V(m_n, w_1) = \alpha$), $u_{m_{2i}}(\mu) \geq 1$ for each $i \in \{1, 2, \ldots, \frac{n-1}{2}\}$ (since $\mu^{(i)}$ has weight $\frac{1}{\alpha}$ in μ and $U(m_{2i}, w_1) = \alpha$), and $v_{w_{2i+1}}(\mu) \geq 1$ for each $i \in \{1, 2, \ldots, \frac{n-1}{2}\}$ (since $\mu^{(i)}$ has weight $\frac{1}{\alpha}$ in μ and $V(m_{2i}, w_{2i+1}) = \alpha$). The welfare of the optimal stable fractional matching must therefore be at least $\mathcal{W}(\mu)$, and thus strictly greater than $(n-1)\alpha$.

We now claim that any ρ-efficient stable fractional matching μ' satisfies $\mu'(m_n, w_1) > 0$. Indeed, assuming otherwise that $\mu'(m_n, w_1) = 0$, the only pair that can give positive utility to man m_1 and woman w_1 is (m_1, w_1). Hence, we must also have $\mu'(m_1, w_1) = 1$, and, as a result, $\mu'(m_{2i}, w_1) = 0$ for all $i \in \{1, 2, \ldots, \frac{n-1}{2}\}$. Then, the only pair that can give positive utility to man m_{2i} and woman w_{2i} is (m_{2i}, w_{2i}), and hence, it must also be that $\mu'(m_{2i}, w_{2i}) = 1$. Consequently, the only pair that can give positive utility to man m_{2i+1} and woman w_{2i+1} is (m_{2i+1}, w_{2i+1}) and, hence, we must have $\mu'(m_{2i+1}, w_{2i+1}) = 1$. The welfare of matching μ' would then be $2n$, which is less than $\rho(n-1)\alpha$ by the assumed bound on α. In other words, the welfare of μ' would be less than ρ times the welfare of the stable fractional matching μ, contradicting the assumption that μ' is ρ-efficient.

The final step in the proof involves showing that for any stable fractional matching μ' with support of size at most $\frac{n-1}{2}\rho$, we must have $\mathcal{W}(\mu') < \rho(n-1)\alpha$; the desired bound on the support size would then follow from the contrapositive. Let $T := \{i \in \{1, 2, \ldots \frac{n-1}{2}\} : \mu'(m_{2i}, w_1) > 0\}$, and $\overline{T} := \{1, 2, \ldots \frac{n-1}{2}\} \setminus T$. Since μ' has support of size at most $\frac{n-1}{2}\rho$ and $\mu'(m_n, w_1) > 0$, it holds that $|T| \leq \frac{n-1}{2}\rho - 1$.

For every $i \in T$, the agents m_{2i}, w_{2i}, m_{2i+1}, and w_{2i+1} can together contribute at most 2α to the welfare. On the other hand, when $i \in \overline{T}$, we have $\mu'(m_{2i}, w_1) = 0$, and the only pair that can give positive utility to man m_{2i} and woman w_{2i} is (m_{2i}, w_{2i}). Therefore, we must have that $\mu'(m_{2i}, w_{2i}) = 1$. Consequently, the only pair that can give positive utility to man m_{2i+1} and woman w_{2i+1} is (m_{2i+1}, w_{2i+1}), and it follows that $\mu'(m_{2i+1}, w_{2i+1}) = 1$. Therefore, when $i \in \overline{T}$, the agents m_{2i}, w_{2i}, m_{2i+1}, and w_{2i+1} can together contribute at most 4 to the welfare. Taking the possible contribution of pair (m_1, w_1) into account, we have that

$$\mathcal{W}(\mu') \leq 2 + 2\alpha|T| + 4|\overline{T}| = 2n + (2\alpha - 4)|T|$$
$$\leq 2n + \rho(n-1)\alpha - 2\alpha - 2(n-1)\rho + 4 < \rho(n-1)\alpha.$$

The equality follows from the definition of $\overline{T}$, the second inequality follows from the bound on $|T|$ above, and the third one from the definition of α. This completes the proof of Theorem [3]. $\square$

Theorem [3] has an interesting algorithmic implication. The fact that the support size can be large even for approximately optimal solutions suggests that a *support enumeration* strategy—which has proven useful in other economic problems [56, 57]—will be strictly more demanding than any of the MILP-based approaches in Section [2.1]. Specifically, a brute force enumeration of all subsets of $\Omega(\rho n)$ matchings requires $\Omega((n!)^{\rho n}) \approx \Omega(n^{\rho n^2})$ time. By contrast, the MILPs (OPT-Stab) and (OPT-Thresh) described in Section [2.1] require $\mathcal{O}(2^{n^2})$ and $\mathcal{O}(n^n)$ time, respectively. A similar comparison can be made for optimal ε-stable fractional matchings. Interestingly, as we will show in Section [4], an ε-stable fractional matching of *small* support can be easily computed, and provides nearly the best approximation ratios achievable by efficient algorithms (under standard complexity-theoretic assumptions).

4. Algorithmic results

We begin the discussion of our algorithmic results with binary valuations. In this setting, OPTIMAL STABLE FRACTIONAL MATCHING reduces to computing a maximum weight matching on a specific weighted graph associated with the given instance.

Theorem 4. *Given an* SMC *instance* $\mathcal{I} = \langle M, W, U, V \rangle$ *with binary valuations, an optimal stable fractional matching for* $\mathcal{I}$ *is, without loss of generality, integral, and can be computed in polynomial time.*

Proof. Let G be the graph representation of $\mathcal{I}$. We assign to each edge (m, w) in G a weight $\gamma(m, w)$, as follows:

$$\gamma(m, w) = \begin{cases} 2 + \frac{1}{n^2} & \text{if } U(m, w) = V(m, w) = 1, \\ 1 & \text{if } U(m, w) = 1, V(m, w) = 0 \text{ or } U(m, w) = 0, V(m, w) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, for any integral matching μ in G, if n_μ denotes the number of agents (men and women) with utility 1 in the SMC instance $\mathcal{I}$, then $n_\mu \leq \sum_{(m,w) \in \mu} \gamma(m, w) < n_\mu + 1$.

Let μ be a maximum weight matching in G. Note that μ can be computed in polynomial time and is, without loss of generality, integral. Also, it follows from the above inequality that μ is an optimal matching for $\mathcal{I}$. We will now argue that μ is stable. Indeed, assuming otherwise, any blocking pair (m, w) must have $0 = u_m(\mu) < U(m, w) = 1$, $0 = v_w(\mu) < V(m, w) = 1$ and $(m, w) \notin \mu$. Thus, if μ contains one or both of the edges (m, w') and (m', w) for some $w' \neq w$ and $m' \neq m$, then we must have that $U(m, w') = 0$ and/or $V(m', w) = 0$. By our definition of weights, this would imply that $\gamma(m, w') = 1$ and/or $\gamma(m', w) = 1$. We can now replace one or both of these edges with the edge (m, w), which has weight $\gamma(m, w) = 2 + 1/n^2$, and obtain a new matching with strictly larger weight—a contradiction. $\square$

The proof of Theorem 4 can be easily extended to the non-bipartite setting (also known as the *stable roommates* problem) to compute an optimal stable fractional matching with binary valuations.

Next, we consider general valuations and show how to exploit stable integral matchings to get an approximate solution for OPTIMAL STABLE FRACTIONAL MATCHING. Let $\sigma_{\max}$ and $\sigma_{\min}$ denote the largest and the smallest non-zero valuation among all agents in $\mathcal{I}$, respectively. We call a man-woman pair (m, w) *light* if either $U(m, w) = 0$ or $V(m, w) = 0$, and *heavy* otherwise. Given an SMC instance $\mathcal{I}$ as input, our algorithm computes a stable integral matching for $\mathcal{I}$, say μ, in two steps: First, it computes a stable integral matching μ_1 using only the heavy pairs (and taking into account the stability constraints in heavy pairs only). Then, it *completes* the solution with a matching μ_2 of maximum welfare using the light pairs subject to feasibility constraints, i.e., using light pairs that do not share any agents with the pairs in μ_1. The light pairs impose no additional constraints on stability, so the resulting matching is stable.

We will show that μ has approximation ratio $1 + \sigma_{\max}/\sigma_{\min}$ (recall that $\sigma_{\max}$ and $\sigma_{\min}$ are the largest and the smallest non-zero valuations, respectively, in the instance $\mathcal{I}$). Let μ^{opt} be an optimal matching for $\mathcal{I}$. Also, let μ_1^{opt} be the set of pairs of μ^{opt} that share an agent with some pair of μ_1, i.e., $\mu_1^{\text{opt}} := \{(m, w) \in \mu^{\text{opt}} : \text{at least one of } m \text{ or } w \text{ is matched in } \mu_1\}$. By definition of μ_2, we have $\mathcal{W}(\mu_2) \geq \mathcal{W}(\mu^{\text{opt}} \setminus \mu_1^{\text{opt}})$. To complete the proof, we will need the following lemma.

Lemma 2. $\mathcal{W}(\mu_1 \setminus \mu_1^{opt}) \geq (1 + \sigma_{\max}/\sigma_{\min})^{-1} \mathcal{W}(\mu_1^{opt} \setminus \mu_1).$

Proof. Our proof constructs a mapping in which every pair $(m, w) \in \mu_1^{\text{opt}} \setminus \mu_1$ is mapped to one of its agents, whom we will call the *witness* of the pair. The mapping is such that the utility of the witness in the matching $\mu_1 \setminus \mu_1^{\text{opt}}$ is at least $(1 + \sigma_{\max}/\sigma_{\min})^{-1} (U(m, w) + V(m, w))$. Note that once we establish the said mapping, the proof will follow, since each agent can be the witness of at most one pair of $\mu_1^{\text{opt}} \setminus \mu_1$ and $\mathcal{W}(\mu_1 \setminus \mu_1^{\text{opt}})$ is at least the total utility of the witnesses in $\mu_1 \setminus \mu_1^{\text{opt}}$.

Consider a light pair $(m, w) \in \mu_1^{\text{opt}} \setminus \mu_1$. The witness is an agent (m or w) who also belongs to a pair of $\mu_1 \setminus \mu_1^{\text{opt}}$; such an agent certainly exists by the definition of μ_1^{opt}. Since all pairs of $\mu_1 \setminus \mu_1^{\text{opt}}$ are heavy, the utility of the witness of (m, w) in $\mu_1 \setminus \mu_1^{\text{opt}}$ is at least $\sigma_{\min} = \frac{\sigma_{\min}}{\sigma_{\max}}(0 + \sigma_{\max}) \geq (1 + \sigma_{\max}/\sigma_{\min})^{-1} (U(m, w) + V(m, w))$, since (m, w) is light.

Now consider a heavy pair $(m, w) \in \mu_1^{\text{opt}} \setminus \mu_1$. If μ_1 contains a pair (m, w') with $U(m, w') \geq U(m, w)$, select agent m to be the witness, otherwise select agent w. Note that in the latter case, stability of μ_1 implies the existence of $(m', w) \in \mu_1$ such that $V(m', w) \geq V(m, w)$. Hence, the utility of the witness of (m, w) in $\mu_1 \setminus \mu_1^{\text{opt}}$ is at least $\min\{U(m, w), V(m, w)\}$, which, in turn, is at least $(1 + \sigma_{\max}/\sigma_{\min})^{-1} (U(m, w) + V(m, w))$. $\square$

Now, Lemma 2 gives the desired approximation ratio, as follows:

$$\begin{aligned} \mathcal{W}(\mu) &= \mathcal{W}(\mu_1) + \mathcal{W}(\mu_2) = \mathcal{W}(\mu_1 \setminus \mu_1^{\text{opt}}) + \mathcal{W}(\mu_1 \cap \mu_1^{\text{opt}}) + \mathcal{W}(\mu_2) \\ &\geq (1 + \sigma_{\max}/\sigma_{\min})^{-1} \mathcal{W}(\mu_1^{\text{opt}} \setminus \mu_1) + \mathcal{W}(\mu_1^{\text{opt}} \cap \mu_1) + \mathcal{W}(\mu^{\text{opt}} \setminus \mu_1^{\text{opt}}) \\ &\geq (1 + \sigma_{\max}/\sigma_{\min})^{-1} \mathcal{W}(\mu^{\text{opt}}). \end{aligned}$$

For ternary valuations in $\{0, 1, \alpha\}$, the above algorithm gives a $(1 + \alpha)$-approximation. An improved approximation for ternary valuations can be achieved using the following modification: When computing the stable integral matching, re-

solve ties in favour of the pairs (m, w) with the highest $U(m, w) + V(m, w)$. The next lemma establishes an improved approximation ratio of $\max\{2, \alpha\}$.

Lemma 3. *The modified algorithm for* SMC *instances with ternary valuations in* $\{0, 1, \alpha\}$ *satisfies* $\mathcal{W}(\mu_1 \setminus \mu_1^{opt}) \geq \min\{\frac{1}{2}, \frac{1}{\alpha}\}\mathcal{W}(\mu_1^{opt} \setminus \mu_1)$.

Proof. For a pair (m, w) of matching $\mu_1^{opt} \setminus \mu_1$, we use the term *neighborhood* to refer to the pairs of $\mu_1 \setminus \mu_1^{opt}$ that use agent m or w. We will show that the total utility from pairs in the neighborhood of (m, w) is at least $\min\{1, 2/\alpha\} (U(m, w) + V(m, w))$. Since each pair of $\mu_1 \setminus \mu_1^{opt}$ can be in the neighborhood of at most two pairs of $\mu_1^{opt} \setminus \mu_1$, this will give us the desired inequality.

Indeed, by the particular way we resolve ties in the ordinal preferences before computing the matching μ_1, a heavy pair (m, w) in $\mu_1^{opt} \setminus \mu_1$ must have a pair of utility at least $U(m, w) + V(m, w)$ in its neighborhood. A light pair (m, w) has $U(m, w) + V(m, w) \leq \alpha$ and certainly has a heavy pair of utility at least 2 in its neighborhood. $\square$

The above discussion is summarized in the following statement.

Theorem 5. *There is a polynomial-time algorithm which, given an* SMC *instance* $\mathcal{I}$ *with an optimal matching* μ^{opt}, *computes a stable integral matching* μ *with* $\mathcal{W}(\mu) \geq \min\{\frac{1}{2}, \frac{1}{\alpha}\}\mathcal{W}(\mu^{opt})$ *if* $\mathcal{I}$ *has ternary valuations in* $\{0, 1, \alpha\}$, *and* $\mathcal{W}(\mu) \geq (1 + \sigma_{\max}/\sigma_{\min})^{-1} \mathcal{W}(\mu^{opt})$ *in general, where* $\sigma_{\max}$ *and* $\sigma_{\min}$ *denote the highest and lowest non-zero valuation in* $\mathcal{I}$, *respectively.*

We conclude this section by considering approximate stability. For general valuations, we present a polynomial-time $1/\varepsilon$-approximation algorithm for OPTIMAL ε-STABLE FRACTIONAL MATCHING, which constructs an ε-stable fractional matching with a small support by combining an optimal matching with a stable integral matching.

Theorem 6. *There is a polynomial-time algorithm that given any* SMC *instance* $\mathcal{I} = \langle M, W, U, V \rangle$ *and any rational* $\varepsilon \in [0, 1]$, *computes a fractional matching* μ *that is* ε-*stable for* $\mathcal{I}$ *such that* $\mathcal{W}(\mu) \geq \varepsilon \mathcal{W}(\mu^{opt})$, *where* μ^{opt} *is an optimal matching for* $\mathcal{I}$.

Proof. Let μ^s be any stable integral matching and μ^{opt} be an optimal matching for $\mathcal{I}$. Note that both μ^s and μ^{opt} can be computed in polynomial time. We will show that $\mu := (1 - \varepsilon)\mu^s + \varepsilon\mu^{opt}$ satisfies the desired properties. Indeed, $\mathcal{W}(\mu) = (1 - \varepsilon)\mathcal{W}(\mu^s) + \varepsilon\mathcal{W}(\mu^{opt}) \geq \varepsilon\mathcal{W}(\mu^{opt})$. Furthermore, since μ^s is stable, we have that for any man-woman pair $(m, w) \in M \times W$, either $u_m(\mu^s) \geq U(m, w)$ or $v_w(\mu^s) \geq V(m, w)$. The former condition implies that $u_m(\mu) \geq (1 - \varepsilon)u_m(\mu^s) \geq (1 - \varepsilon)U(m, w)$, while the latter condition gives $v_w(\mu) \geq (1 - \varepsilon)V(m, w)$. Either way, the pair (m, w) is ε-stable. $\square$

In particular, Theorem 6 shows that a $\frac{1}{2}$-stable fractional matching with welfare at least half of that of an optimal fractional matching (and therefore, that of an optimal stable fractional matching) can be computed in polynomial time. In Appendix D, we provide a slightly stronger welfare guarantee: There is a polynomial-time algorithm that computes a $\frac{1}{2}$-stable fractional matching with welfare at least that of an optimal (exactly) stable fractional matching.

5. Hardness of approximation

In this section, we present our inapproximability statements, which are by far the technically most involved results in the paper. We present polynomial-time reductions which, given a 3SAT formula ϕ of a particular structure, construct SMC instances that simulate the evaluation of ϕ for every variable assignment. The constructed SMC instances consist of several gadgets including an *accumulator*. The simulation of the evaluation of ϕ by the SMC instance is such that:

(a) when ϕ has a satisfying assignment, there is a stable (or ε-stable) fractional matching where the contribution of the agents in the accumulator gadget to the welfare can be large and dominates the contribution from the remaining SMC instance, and

(b) when ϕ is not satisfiable, the contribution of the accumulator and, subsequently, the total welfare of any stable (or ε-stable) fractional matching is very small.

Hence, distinguishing between SMC instances with stable (or ε-stable) fractional matchings of very high and very low welfare would allow us to decide 3SAT. We have two inapproximability statements: Theorem 7 for OPTIMAL STABLE FRACTIONAL MATCHING and Theorem 8 for OPTIMAL ε-STABLE FRACTIONAL MATCHING.

Theorem 7. *For every constant* $\delta > 0$, *it is NP-hard to approximate* OPTIMAL STABLE FRACTIONAL MATCHING *for* SMC *instances with ternary valuations in* $\{0, 1, \alpha\}$ *to within a factor of (i)* $\alpha - 1/2 - \delta$ *if* $\alpha = \mathcal{O}(n)$, *and (ii)* $\Omega(n^{1-\delta})$ *otherwise.*

Theorem 8. *For any constant* $\varepsilon \in (0, 0.03]$ *and* $\delta > 0$, *it is NP-hard to approximate* OPTIMAL ε-STABLE FRACTIONAL MATCHING *to within a factor of* $1/\varepsilon - \delta$.
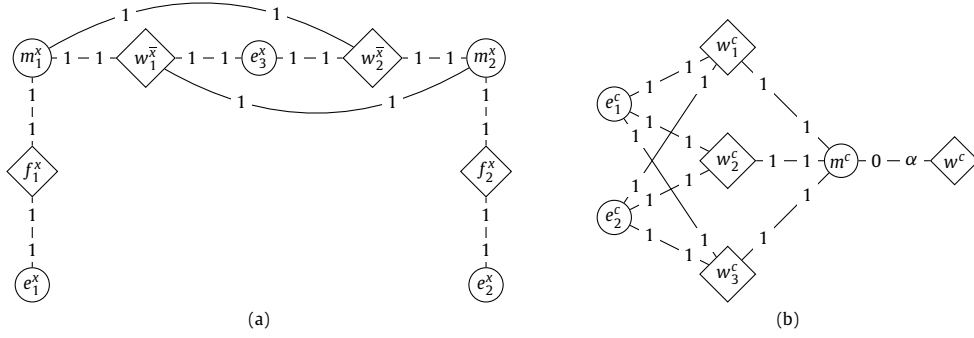
Fig. 5. (a) The variable gadget corresponding to the variable x. (b) The clause gadget corresponding to the clause c and its CA-connector (m^c, w^c).

We will prove Theorem 7 here; the proof of Theorem 8, which uses similar gadgets but is slightly more involved, appears in Appendix C. Since the proof is long, we have divided it into three parts: the description of the reduction (Section 5.1), technical claims with gadget properties (Section 5.2), and the proof of the inapproximability result (Section 5.3).

5.1. The reduction

In particular, we present a polynomial-time reduction from 2P2N-3SAT, the special case of 3SAT consisting of 3-CNF clauses in which every variable appears four times: twice as a positive literal and twice as a negative one. 2P2N-3SAT is known to be NP-hard [58]. Our reduction takes as input an instance of 2P2N-3SAT consisting of N (boolean) variables $x_1, x_2, \ldots, x_N$, and a 3-CNF formula ϕ with $L = 4N/3$ clauses $c_1, c_2, \ldots, c_L$. Without loss of generality, we assume that each clause in ϕ consists of distinct literals.

Given the instance of 2P2N-3SAT, our reduction generates an instance $\mathcal{I} = \langle M, W, U, V \rangle$ of Optimal Stable Fractional Matching. As usual, we denote by n the number of men (or women) in $\mathcal{I}$. We will use a positive integer parameter k which will determine the size of n; in particular, $n = \mathcal{O}(N + k)$. We define $\mathcal{I}$ by referring to its graph representation, which consists of *variable gadgets*, *clause gadgets*, *variable-clause connectors*, an *accumulator*, and *clause-accumulator connectors*. For each gadget, we classify the edges (i.e., man-woman pairs and their valuations) into the following three types:

- *man-heavy* edges (m, w) with $U(m, w) = \alpha$ and $V(m, w) = 0$,
- *woman-heavy* edges (m, w) with $U(m, w) = 0$ and $V(m, w) = \alpha$, and
- *balanced* edges (m, w) with $U(m, w) = V(m, w) = 1$.

Recall that any pair (m, w) that does not appear as an edge in the graph representation has $U(m, w) = V(m, w) = 0$.

The instance $\mathcal{I}$ has a variable gadget for every variable x, which consists of five men $m_1^x, m_2^x, e_1^x, e_2^x, e_3^x$, four women $w_1^{\bar{x}}$, $w_2^{\bar{x}}, f_1^x, f_2^x$ and the ten balanced edges $(e_1^x, f_1^x), (m_1^x, f_1^x), (m_1^x, w_1^{\bar{x}}), (e_3^x, w_1^{\bar{x}}), (e_3^x, w_2^{\bar{x}}), (m_2^x, w_2^{\bar{x}}), (m_2^x, f_2^x), (e_2^x, f_2^x), (m_1^x, w_2^{\bar{x}})$, and $(m_2^x, w_1^{\bar{x}})$, as shown in Fig. 5a.

For every clause c, instance $\mathcal{I}$ has a clause gadget with three men m^c, e_1^c, e_2^c, three women w_1^c, w_2^c, w_3^c, and the nine balanced edges between them, as shown in Fig. 5b.

For every appearance of a literal in a clause, there is a variable-clause connector (or *VC-connector*). The structure of VC-connectors depends on whether they correspond to positive or negative literals and the value of α. In each case, we identify one edge of the VC-connector as the *input*, and either one or two edges as the *output*.

Specifically, for every positive literal x whose i-th appearance ($i \in \{1, 2\}$) is as the j-th literal ($j \in \{1, 2, 3\}$) of clause c, $\mathcal{I}$ has a VC-connector defined as follows:

- When $\alpha \geq 2$, the VC-connector consists of a single woman-heavy edge between m_i^x (from the variable gadget corresponding to variable x) and w_j^c (from the clause gadget corresponding to clause c), as shown in Fig. 6a. This edge is simultaneously the input and the output edge of the VC-connector.
- When $\alpha \in (3/2, 2)$, the VC-connector consists of woman $w^{x,c}$, man $m^{x,c}$, the woman-heavy edges $(m_i^x, w^{x,c})$ and $(m^{x,c}, w_j^c)$, and the balanced edge $(m^{x,c}, w^{x,c})$, as shown in Fig. 6b. Here, $(m_i^x, w^{x,c})$ is the input and $(m^{x,c}, w_j^c)$ is the output edge.

For every negative literal $\bar{x}$ whose i-th appearance ($i \in \{1, 2\}$) is as the j-th literal ($j \in \{1, 2, 3\}$) of clause c, $\mathcal{I}$ has a VC-connector defined as follows:
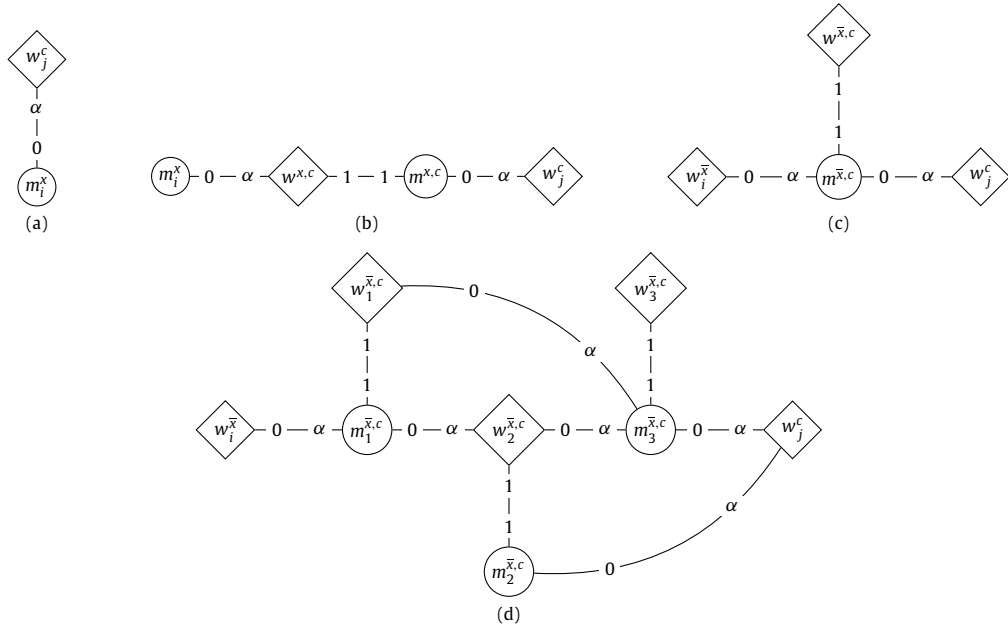
Fig. 6. VC-connectors corresponding to clause c and positive literal x for (a) $\alpha \geq 2$ and (b) $\alpha \in (3/2, 2)$, and to clause c and negative literal $\bar{x}$ for (c) $\alpha \geq 2$ and (d) $\alpha \in (3/2, 2)$.

- When $\alpha \geq 2$, the VC-connector consists of man $m^{\bar{x},c}$, woman $w^{\bar{x},c}$, the man-heavy edge $(m^{\bar{x},c}, w_i^{\bar{x}})$, the balanced edge $(m^{\bar{x},c}, w^{\bar{x},c})$, and the woman-heavy edge $(m^{\bar{x},c}, w_j^c)$, as shown in Fig. 6c. Here, $(m^{\bar{x},c}, w_i^{\bar{x}})$ is the input and $(m^{\bar{x},c}, w_j^c)$ is the output edge.

- When $\alpha \in (3/2, 2)$, the VC-connector consists of three men $m_1^{\bar{x},c}, m_2^{\bar{x},c}, m_3^{\bar{x},c}$, three women $w_1^{\bar{x},c}, w_2^{\bar{x},c}, w_3^{\bar{x},c}$, the man-heavy edges $(m_1^{\bar{x},c}, w_i^{\bar{x}}), (m_3^{\bar{x},c}, w_1^{\bar{x},c}), (m_3^{\bar{x},c}, w_2^{\bar{x},c})$, the woman-heavy edges $(m_1^{\bar{x},c}, w_2^{\bar{x},c}), (m_2^{\bar{x},c}, w_j^c), (m_3^{\bar{x},c}, w_j^c)$, and the balanced edges $(m_1^{\bar{x},c}, w_1^{\bar{x},c}), (m_2^{\bar{x},c}, w_2^{\bar{x},c}), (m_3^{\bar{x},c}, w_3^{\bar{x},c})$, as shown in Fig. 6d. In this case, the VC-connector has one input edge $(m_1^{\bar{x},c}, w_i^{\bar{x}})$ and two output edges $(m_2^{\bar{x},c}, w_j^c)$ and $(m_3^{\bar{x},c}, w_j^c)$.
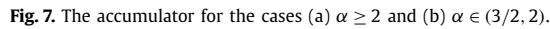
The accumulator (Fig. 7) of instance $\mathcal{I}$ has different structure depending on the value of α. Its size depends on the positive integer parameter k.

- When $\alpha \geq 2$ (see Fig. 7a), the accumulator consists of man m_i and woman w_i for all $i \in \{1, \ldots, k\}$, men e_i^1 and e_i^2 and woman f_i^1 for all $i \in \{1, \ldots, k-1\}$, man e_i^3 and women f_i^2 and f_i^3 for all $i \in \{2, \ldots, k\}$, and woman w^c for every clause c of ϕ. The edges comprise of man-heavy edges (m_i, w_{i-1}) and (e_i^3, f_i^2) for all $i \in \{2, \ldots, k\}$ and (e_i^2, w_i) for all $i \in \{1, \ldots, k-1\}$, balanced edges (m_1, w^c) for every clause c, which we call *tine* edges, (e_i^1, w_i) for all $i \in \{1, \ldots, k-1\}$, and (m_i, f_i^2) for all $i \in \{2, \ldots, k\}$, and woman-heavy edges (m_i, w_i) for all $i \in \{1, \ldots, k\}$, (e_i^1, f_i^1) for all $i \in \{1, \ldots, k-1\}$, and (m_i, f_i^3) for all $i \in \{2, \ldots, k\}$.

- When $\alpha \in (3/2, 2)$ (see Fig. 7b), the accumulator consists of man m_i, woman w_i for all $i \in \{1, \ldots, k\}$, man e_i^1 and woman f_i^1 for all $i \in \{1, \ldots, k-1\}$, man e_i^2 and woman f_i^2 for all $i \in \{2, \ldots, k\}$, and woman w^c for every clause c of ϕ. The edges comprise of man-heavy edges (m_i, w_{i-1}) and (e_i^2, f_i^2) for all $i \in \{2, \ldots, k\}$ and (m_i, f_{i-1}^2) for all $i \in \{3, \ldots, k\}$, the balanced edges (m_1, w^c) for every clause c (tine edges), (e_i^1, w_i) for all $i \in \{1, \ldots, k-1\}$ and (m_i, f_i^2) for all $i \in \{2, \ldots, k\}$, and the woman-heavy edges (m_i, w_i) for all $i \in \{1, \ldots, k\}$, and (e_i^1, f_i^1) and (e_i^1, w_{i+1}) for all $i \in \{1, \ldots, k-1\}$.

Finally, instance $\mathcal{I}$ has a clause-accumulator connector (or *CA-connector*) for every clause c of ϕ consisting of the woman-heavy edge (m^c, w^c) between the man m^c (from the clause gadget corresponding to clause c) and woman w^c (from the accumulator); see Fig. 5b. Notice that the above construction has more women than men. To restore balance, we pad the instance with extra (isolated) men that neither value nor are valued by any other agent. This completes the construction of the reduced instance.

5.2. Gadget properties

We will now prove several important properties (Claims 1-4) of our construction.

Fig. 7. The accumulator for the cases (a) $\alpha \geq 2$ and (b) $\alpha \in (3/2, 2)$.

Claim 1. *For every variable x, a stable fractional matching μ satisfies at least one of the following:*

(1) $\mu(m_1^x, w_1^{\overline{x}}) + \mu(m_1^x, w_2^{\overline{x}}) + \mu(m_1^x, f_1^x) = 1$ *and* $\mu(m_2^x, w_1^x) + \mu(m_2^x, w_2^x) + \mu(m_2^x, f_2^x) = 1$.
(2) $\mu(m_1^x, w_1^{\overline{x}}) + \mu(m_2^x, w_1^{\overline{x}}) + \mu(e_3^x, w_1^{\overline{x}}) = 1$ *and* $\mu(m_1^x, w_2^{\overline{x}}) + \mu(m_2^x, w_2^{\overline{x}}) + \mu(e_3^x, w_2^{\overline{x}}) = 1$.

Proof. Suppose, for contradiction, that for some $i, j \in \{1, 2\}$, we have $\mu(m_i^x, w_1^{\overline{x}}) + \mu(m_i^x, w_2^{\overline{x}}) + \mu(m_i^x, f_i^x) < 1$ and $\mu(m_1^x, w_j^{\overline{x}}) + \mu(m_2^x, w_j^{\overline{x}}) + \mu(e_3^x, w_j^{\overline{x}}) < 1$. Then, both m_i^x and $w_j^{\overline{x}}$ will have utility strictly less than 1 in μ, and thus the pair $(m_i^x, w_j^{\overline{x}})$ would be blocking. □

We remark that the two conditions in the statement of Claim 1 affect the weight of the input edges of the VC-connectors that are attached to the variable gadget in any stable fractional matching. In particular, condition (1) implies that the weight assigned to the input edges of the VC-connectors that correspond to the two appearances of the positive literal x in clauses must be 0. To see why, observe that these input edges are incident to nodes m_1^x and m_2^x, and the total weight of all edges

incident to each of these nodes cannot exceed 1. Condition (2) has a similar implication for the edges associated with the negative literal $\overline{x}$.

Claim 2. *Any stable fractional matching that assigns a weight of 0 to the input edge of a VC-connector must assign a weight of 0 to its output edge(s) as well.*

Proof. For $\alpha \geq 2$, the claim holds trivially for VC-connectors corresponding to positive literals (Fig. 6a). Consider a VC-connector corresponding to a negative literal $\overline{x}$ and a clause c containing it (Fig. 6c). Observe that, besides the input edge $(m^{\overline{x},c}, w^{\overline{x}}_i)$, the edge $(m^{\overline{x},c}, w^{\overline{x},c})$ is the only balanced or man-heavy edge that is incident to man $m^{\overline{x},c}$ and the only balanced (or woman-heavy) edge incident to woman $w^{\overline{x},c}$. Hence, stability of the edge $(m^{\overline{x},c}, w^{\overline{x},c})$ requires a weight of 1 assigned to it when the weight assigned to input edge $(m^{\overline{x},c}, w^{\overline{x}}_i)$ is 0. Then, the output edge $(m^{\overline{x},c}, w^c_j)$, which is also incident to the node $m^{\overline{x},c}$, must have a weight of 0 as well.

We now consider the case $\alpha \in (3/2, 2)$. First consider a VC-connector corresponding to a positive literal x and a clause c containing it (Fig. 6b). The edge $(m^{x,c}, w^{x,c})$ is the only balanced or man-heavy edge that is incident to man $m^{x,c}$ and, besides the input edge $(m^x_i, w^{x,c})$, the only balanced (or woman-heavy) edge incident to woman $w^{x,c}$. Hence, stability of edge $(m^{x,c}, w^{x,c})$ requires a weight of 1 assigned to it when the weight assigned to edge $(m^x_i, w^{x,c})$ is 0. Then, the output edge $(m^{x,c}, w^c_j)$, which is also incident to node $m^{x,c}$, must have a weight of 0 as well.

Finally, consider a VC-connector corresponding to a negative literal $\overline{x}$ and a clause c containing it (Fig. 6d). Observe that $(m^{\overline{x},c}_1, w^{\overline{x},c}_1)$ is the only balanced or woman-heavy edge that is incident to woman $w^{\overline{x},c}_1$ and, besides the input edge $(m^{\overline{x},c}_1, w^{\overline{x}}_i)$, the only balanced or man-heavy edge incident to man $m^{\overline{x},c}_1$. Also, $(m^{\overline{x},c}_2, w^{\overline{x},c}_2)$ is the only balanced or man-heavy edge that is incident to man $m^{\overline{x},c}_2$ and, besides edge $(m^{\overline{x},c}_1, w^{\overline{x},c}_2)$, the only balanced or woman-heavy edge to woman $w^{\overline{x},c}_2$. Furthermore, $(m^{\overline{x},c}_3, w^{\overline{x},c}_3)$ is the only balanced or man-heavy edge that is incident to woman $w^{\overline{x},c}_3$ and, besides edges $(m^{\overline{x},c}_3, w^{\overline{x},c}_1)$ and $(m^{\overline{x},c}_3, w^{\overline{x},c}_2)$, the only balanced or man-heavy edge incident to man $m^{\overline{x},c}_3$.

Hence, stability of edge $(m^{\overline{x},c}_1, w^{\overline{x},c}_1)$ requires a weight of 1 assigned to it when the weight assigned to edge $(m^{\overline{x},c}_1, w^{\overline{x}}_i)$ is 0. Then, edges $(m^{\overline{x},c}_1, w^{\overline{x},c}_2)$ and $(m^{\overline{x},c}_3, w^{\overline{x},c}_1)$ must have a weight of 0. Then, stability of edge $(m^{\overline{x},c}_2, w^{\overline{x},c}_2)$ requires a weight of 1 assigned to it and the edge $(m^{\overline{x},c}_3, w^{\overline{x},c}_2)$ and the output edge $(m^{\overline{x},c}_2, w^c_j)$ must have a weight of 0. Then, stability of edge $(m^{\overline{x},c}_3, w^{\overline{x},c}_3)$ requires a weight of 1 assigned to it. Hence, the output edge $(m^{\overline{x},c}_3, w^c_j)$ must have a weight of 0 as well. □

Claim 3. *Any stable fractional matching that assigns a weight of 0 to all output edges of the VC-connectors of clause c must assign a weight of 0 to the CA-connector of clause c as well.*

Proof. Let ℓ_1, ℓ_2, and ℓ_3 be the literals of clause c. Consider, for the sake of contradiction, a stable fractional matching that assigns (1) a weight of 0 to all output edges of the VC-connectors corresponding to literals ℓ_i and clause c, and (2) strictly positive weight to edge (m^c, w^c) of the CA-connector for clause c. Note that condition (2) implies that the total weight on the edges (m^c, w^c_1), (m^c, w^c_2) and (m^c, w^c_3) is strictly smaller than 1. Since these are the only balanced or man-heavy edges incident to man m^c, the stability of these edges is guaranteed by a utility of (at least) 1 for each of the agents w^c_1, w^c_2, and w^c_3.

Note that besides the output edges of the VC-connectors, the edges (e^c_1, w^c_i), (e^c_2, w^c_i), and (m^c, w^c_i) are the only balanced or woman-heavy edges incident to agent w^c_i for all $i \in \{1, 2, 3\}$. Along with condition (1), this implies that the weight assigned to these three edges is at least 1. Hence, the total weight on the nine edges of the clause gadget is at least 3, i.e., strictly more than 2 for the six edges incident to men e^c_1 and e^c_2, violating the definition of a fractional matching. □

Claim 4. *Any stable fractional matching that assigns a weight of 0 to some CA-connector must assign a total weight of 1 to the tine edges and a weight of 1 to every balanced edge of the accumulator.*

Proof. Assume that a weight of 0 has been assigned to the edge $(m^{c'}, w^{c'})$ of the CA-connector corresponding to some clause c'. Since this is the only woman-heavy edge that is incident to agent $w^{c'}$ and there is no man-heavy edge incident to agent m_1, stability on the edge $(m_1, w^{c'})$ requires that the total weight of the tine edges (m_1, w^c) (for every clause c) is (at least) 1. Hence, the weight of the edge (m_1, w_1) is 0. We will complete the proof by distinguishing between the two different accumulator structures, depending on whether $\alpha \geq 2$ or $\alpha \in (3/2, 2)$.

When $\alpha \geq 2$, it suffices to show that for all $i \in \{1, \ldots, k-1\}$, if the weight of edge (m_i, w_i) is 0, then the weight of the balanced edges (e^1_i, w_i) and (m_{i+1}, f^2_{i+1}) is 1 and the weight of edge (m_{i+1}, w_{i+1}) is 0. Indeed, observe that, edge (e^1_i, w_i) is the only balanced or man-heavy edge incident to man e^1_i and, besides edge (m_i, w_i), the only balanced or woman-heavy edge incident to woman w_i. Hence, the balanced edge (e^1_i, w_i) must have a weight of 1 and the edge (m_{i+1}, w_i) a weight of 0.

Then, edge (m_{i+1}, f_{i+1}^2) is the only balanced or woman-heavy edge incident to woman f_{i+1}^2 and, besides edge (m_{i+1}, w_i), the only balanced or man-heavy edge incident to man m_{i+1}. Hence, the balanced edge (m_{i+1}, f_{i+1}^2) must have a weight of 1 and the edge (m_{i+1}, w_{i+1}) a weight of 0.

When $\alpha \in (3/2, 2)$, it suffices to show that for all $i \in \{1, \ldots, k-1\}$, if the weight of the edge (m_i, w_i) (and, if they exist, the edges (e_{i-1}^1, w_i) and (m_{i+1}, f_i^2)) is 0, then the weight of the balanced edges (e_i^1, w_i) and (m_{i+1}, f_{i+1}^2) is 1, and the weight of the edges (m_{i+1}, w_{i+1}) and (e_i^1, w_{i+1}) (and, if it exists, the edge (m_{i+2}, f_{i+1}^2)) is 0. Indeed, observe that the edge (e_i^1, w_i) is the only balanced or man-heavy edge incident to man e_i^1. Furthermore, in addition to the edge (m_i, w_i) (and, if it exists, the edge (e_{i-1}^1, w_i)), it is the only balanced or woman-heavy edge incident to woman w_i. Hence, the balanced edge (e_i^1, w_i) must have a weight of 1, and the edge (m_{i+1}, w_i) (and, if it exists, the edge (e_i^1, w_{i+1})) must have a weight of 0. Then, the edge (m_{i+1}, f_{i+1}^2) is, besides (m_{i+1}, f_i^2) and (m_{i+1}, w_i), the only balanced or man-heavy edge incident to man m_{i+1}, and the only balanced or woman-heavy edge incident to woman f_{i+1}^2. Hence, the balanced edge (m_{i+1}, f_{i+1}^2) must have a weight of 1, and the edge (m_{i+1}, w_{i+1}) (and, if it exists, the edge (m_{i+2}, f_{i+1}^2)) must have a weight of 0. □

5.3. Proof of inapproximability

Lemma 4. *If formula ϕ is not satisfiable, then any stable fractional matching of $\mathcal{I}$ has welfare at most $80\alpha N + 4(k-1)$.*

Proof. We will first show that if ϕ is not satisfiable, then any stable fractional matching of $\mathcal{I}$ assigns weight 0 to some CA-connector. For the sake of contradiction, consider a stable fractional matching that assigns a strictly positive weight to all CA-connectors. We will construct a truth assignment for the formula ϕ (contradicting the assumption of the lemma) by repeating the following process for every clause c of ϕ: Let ℓ be a literal that appears in c such that the output edge(s) of the VC-connector, that corresponds to the appearance of ℓ in c, have strictly positive total weight. By Claim 3, such a literal must exist. We set ℓ to 1 (true). For every variable that has not received a value in this way, we arbitrarily set it to 1.

The above assignment satisfies all the clauses. To show that it is also valid, we need to argue that there is no variable x such that both literals x and $\overline{x}$ have been set to 1. Assume, to the contrary, that literal x is set to 1 due to its appearance in a clause c_1, and literal $\overline{x}$ is set to 1 due to its appearance in a different clause c_2. Thus, in the above assignment, the output edge(s) of the VC-connector between the literal x and the clause c_1, as well as the VC-connector between the literal $\overline{x}$ and the clause c_2 have strictly positive (total) weight. By Claim 2, the input edges of both VC-connectors also have strictly positive weight. Let $i_1, i_2 \in \{1, 2\}$ be such that the i_1-th appearance of x is in the clause c_1 and the i_2-th appearance of $\overline{x}$ is in the clause c_2. Therefore, the said input edges are incident to the nodes $m_{i_1}^x$ and $w_{i_2}^{\overline{x}}$. Using Claim 1, we get that the total weight on the edges incident to one of $m_{i_1}^x$ or $w_{i_2}^{\overline{x}}$ exceeds 1, contradicting feasibility. Thus, the above assignment must be valid, which, in turn, implies that any stable fractional matching assigns weight 0 to some CA-connector.

By Claim 4, the contribution of the accumulator to the welfare is exactly $4k-2$ (2 from the tine edges plus 2 from each balanced edge). Let us now consider the contribution of the edges that do not belong to the accumulator. This comprises of

- a total value of 20 for the ten balanced edges of each of the N variable gadgets,
- a total value of α (respectively, $2+2\alpha$) for the edges of each of the $2N$ VC-connectors corresponding to a positive literal when $\alpha \geq 2$ (respectively, $\alpha \in (3/2, 2)$),
- a total value of $2+2\alpha$ (respectively, $6+6\alpha$) for the edges of each of the $2N$ VC-connectors corresponding to a negative literal when $\alpha \geq 2$ (respectively, $\alpha \in (3/2, 2)$),
- a total value of $18 + \alpha$ for the nine balanced edges of each of the $4N/3$ clause gadgets and their corresponding CA-connectors.

It can be easily seen that $80\alpha N - 2$ is a (loose) upper bound on the total value from these edges. □

Lemma 5. *If ϕ is satisfiable, then there exists a stable fractional matching of $\mathcal{I}$ with welfare at least $4(k-1)(\alpha - 1/2)$.*

Proof. Starting from a satisfying assignment for ϕ, we will construct a stable fractional matching μ in which the welfare of the accumulator gadget is at least $4(k-1)(\alpha - 1/2)$.

Variable gadgets. For the edges of the variable gadget of the variable x, μ is defined as:

- If x is true, then $\mu(m_1^x, w_1^{\overline{x}}) = \mu(e_3^x, w_1^{\overline{x}}) = \mu(e_3^x, w_2^{\overline{x}}) = \mu(m_2^x, w_2^{\overline{x}}) = 1/2$, $\mu(e_1^x, f_1^x) = \mu(e_2^x, f_2^x) = 1$, and the remaining edges have weight 0.
- If x is false, then $\mu(e_3^x, w_1^{\overline{x}}) = \mu(e_3^x, w_2^{\overline{x}}) = 1/2$, $\mu(m_1^x, f_1^x) = \mu(m_2^x, f_2^x) = 1$, and the remaining edges have weight 0.

Clause gadgets and CA-connectors. For each clause, select one of the true literals (tie-break arbitrarily) and call it *active*. Note that each clause has an active literal in a satisfying assignment. Consider the clause c, and let ℓ_i be its active literal for some

$i \in \{1, 2, 3\}$. Also, let $i_1, i_2 \in \{1, 2, 3\} \setminus \{i\}$ denote the other two indices. Set $\mu(e_1^c, w_{i_1}^c) = \mu(e_2^c, w_{i_2}^c) = 1$, and set the weight of the remaining balanced edges to 0. Assign a weight of 1 to the CA-connector, i.e., $\mu(m^c, w^c) = 1$.

VC-connectors. For every non-active VC-connector, set the weight of its balanced edges (if any) to 1 and the weight of the remaining edges to 0. For every active VC-connector corresponding to the i-th appearance of the positive literal x as the j-th literal of clause c ($i \in \{1, 2\}$, $j \in \{1, 2, 3\}$), the weights of its edges are as follows:

- When $\alpha \geq 2$, we set $\mu(m_i^x, w_j^c) = 1/2$.
- When $\alpha \in (3/2, 2)$, we set $\mu(m_i^x, w^{x,c}) = 1/2$, $\mu(m^{x,c}, w^{x,c}) = 1 - \alpha/2$, and $\mu(m^{x,c}, w_j^c) = 1/\alpha$.

For every active VC-connector corresponding to the i-th appearance of the negative literal x as the j-th literal of clause c ($i \in \{1, 2\}$, $j \in \{1, 2, 3\}$), the weights of its edges are as follows:

- When $\alpha \geq 2$, we set $\mu(m^{\overline{x},c}, w_i^{\overline{x}}) = \mu(m^{\overline{x},c}, w_j^c) = 1/2$ and $\mu(m^{\overline{x},c}, w^{\overline{x},c}) = 0$.
- When $\alpha \in (3/2, 2)$, we set $\mu(m_1^{\overline{x},c}, w_i^{\overline{x}}) = 1/2$, $\mu(m_1^{\overline{x},c}, w_1^{\overline{x},c}) = 1 - \alpha/2$, $\mu(m_1^{\overline{x},c}, w_2^{\overline{x},c}) = (\alpha - 1)/2$, $\mu(m_2^{\overline{x},c}, w_2^{\overline{x},c}) = 1 - (\alpha^2 - \alpha)/2$, $\mu(m_2^{\overline{x},c}, w_j^c) = 2/\alpha - 1$, $\mu(m_3^{\overline{x},c}, w_1^{\overline{x},c}) = 1/\alpha$, $\mu(m_3^{\overline{x},c}, w_2^{\overline{x},c}) = \mu(m_3^{\overline{x},c}, w_3^{\overline{x},c}) = 0$, $\mu(m_3^{\overline{x},c}, w_j^c) = 1 - 1/\alpha$.

Accumulator. We set $\mu(m_1, w^c) = 0$ for every tine edge (m_1, w^c) of the accumulator. Furthermore:

- When $\alpha \geq 2$, we set $\mu(m_i, w_i) = 1/\alpha$ for all $i \in \{1, \ldots, k\}$, $\mu(e_i^2, w_i) = 1 - 2/\alpha$, $\mu(m_{i+1}, w_i) = 1/\alpha$, $\mu(e_i^1, f_i^1) = 1$, $\mu(e_i^1, w_i) = 0$ for all $i \in \{1, \ldots, k-1\}$, $\mu(m_i, f_i^2) = 0$, $\mu(m_i, f_i^3) = 1 - 2/\alpha$, and $\mu(e_i^3, f_i^2) = 1$ for all $i \in \{2, \ldots, k\}$. Among these, any edge with a positive weight is either man- or woman-heavy, and hence, its contribution to the social welfare is α times its weight. It can be verified that the total contribution is $4(k-1)(\alpha - 1/2) + 1$.
- When $\alpha \in (3/2, 2)$, we set $\mu(m_1, w_1) = 1/\alpha$, $\mu(m_2, w_2) = \alpha + 1/\alpha - 2$, $\mu(m_i, w_i) = 1 - 1/\alpha$ for all $i \in \{3, \ldots, k\}$, $\mu(m_{i+1}, w_i) = 1 - 1/\alpha$ for all $i \in \{1, \ldots, k-1\}$, $\mu(e_i^1, w_i) = 0$ for all $i \in \{1, \ldots, k-1\}$, $\mu(m_2, f_2^2) = 2 - \alpha$, $\mu(m_i, f_i^2) = 0$ for all $i \in \{3, \ldots, k\}$, $\mu(e_1^1, f_1^1) = \alpha - 1$, $\mu(e_2^2, f_2^2) = \alpha - 2/\alpha$, $\mu(e_k^2, f_k^2) = 1$, $\mu(e_i^1, f_i^1) = 2 - 2/\alpha$ for all $i \in \{2, \ldots, k-1\}$, $\mu(e_i^2, f_i^2) = 2 - 2/\alpha$ for all $i \in \{3, \ldots, k-1\}$, $\mu(e_1^1, w_2) = 2 - \alpha$, $\mu(e_i^1, w_{i+1}) = 2/\alpha - 1$ for all $i \in \{2, \ldots, k-1\}$, and $\mu(m_{i+1}, f_i^2) = 2/\alpha - 1$ for all $i \in \{2, \ldots, k-1\}$. Except for the balanced edge (m_2, f_2^2), every edge with a positive weight among the ones listed above is either man- or woman-heavy, and hence, its contribution to the social welfare is α times its weight. It can be verified that the total contribution in this case is $4(k-1)(\alpha - 1/2) + 2\alpha^2 - 7\alpha + 7$.

In each case, the accumulator contributes at least $4(k-1)(\alpha - 1/2)$ to the social welfare, as desired.

The feasibility of μ can be verified by inspection. To see why μ is stable, note that we only need to check for the balanced edges, as the man- or woman-heavy edges and the remaining pairs do not impose any constraints on stability. For the balanced edges, stability is established by the following series of observations (we will use the term 'stabilized by' to denote that an agent's utility is at least 1): The variable gadget for the variable x (Fig. 5a) is stabilized by the agents f_1^x, f_2^x, e_3^x along with m_1^x, m_2^x (if x is true) or $w_1^{\overline{x}}$, $w_2^{\overline{x}}$ (if x is false). The clause gadget for clause c (Fig. 5b) with active index i (and non-active indices i_1 and i_2) is stabilized by the agents e_1^c, e_2^c, w_i^c, $w_{i_1}^c$, $w_{i_2}^c$; in particular, the edge (m^c, w_i^c) is stabilized by w_i^c because an active literal triggers the woman-heavy edge in the VC-connector. A VC connector is stabilized by $w^{x,c}$ (Fig. 6b), $m^{\overline{x},c}$ (Fig. 6c), or $m_1^{\overline{x},c}$, $w_2^{\overline{x},c}$, and $m_3^{\overline{x},c}$ (Fig. 6d). Finally, the tine edges in the accumulator (Fig. 7) are stabilized by $w^{c_1}, \ldots, w^{c_L}$ (because we trigger the CA-connector), and the remaining balanced edges are stabilized by w_i's and m_i's except for m_1. Overall, μ is a feasible stable fractional matching. $\square$

We are ready to prove Theorem 7. If $\alpha < N^{1+1/\delta}$, we use our construction with any k satisfying $k - 1 \geq \frac{20\alpha N(\alpha - 1/2 - \delta)}{\delta}$. It is easy to verify that the reduction is polynomial-time. Furthermore, from Lemma 4, we know that the welfare of μ when ϕ is not satisfiable is at most

$$80\alpha N + 4(k-1) \leq \frac{4(k-1)\delta}{\alpha - 1/2 - \delta} + 4(k-1) = \frac{4(k-1)(\alpha - 1/2)}{\alpha - 1/2 - \delta}.$$

This number is at least $\alpha - 1/2 - \delta$ times smaller than the welfare of μ when ϕ is satisfiable (Lemma 5). This establishes the inapproximability bound in part (i) of Theorem 7.

If $\alpha \geq N^{1+1/\delta}$, we use our construction with $k = N^{1+1/\delta}$. Once again, the reduction is polynomial-time, and the instance $\mathcal{I}$ has $n = \Theta(N^{1+1/\delta})$ men and women. Observe that $\alpha = \Omega(n)$, $k = \Theta(n)$, and $N = \mathcal{O}(n^{\delta})$. Hence, the welfare of μ when ϕ is not satisfiable is at most

$$80\alpha N + 4(k-1) \leq 80\alpha N + 4N^{1+1/\delta} \leq 84\alpha N = \mathcal{O}(\alpha n^{\delta}).$$

On the other hand, the welfare of μ when ϕ is satisfiable is at least $4(k-1)(\alpha - 1/2)$, i.e., $\Omega(\alpha n)$. This establishes the bound in part (ii), and with it, completes the proof of Theorem 7.
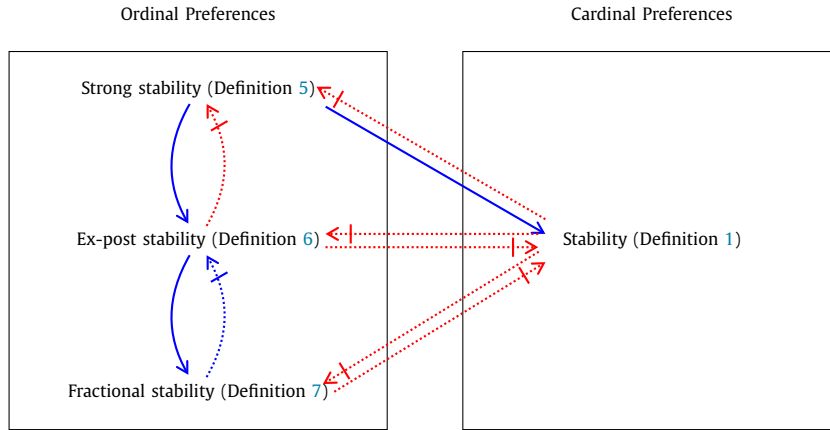
Fig. 8. Relationship between stability notions for fractional matchings under ordinal and cardinal preferences. A solid (respectively, dotted) arrow indicates that the implication holds (respectively, does not hold) in that direction. The implications for strict and weak preferences are denoted by red and blue arrows, respectively. (For interpretation of the colors in the figure, the reader is referred to the web version of this article.)

6. Concluding remarks

We studied stable fractional matchings in a cardinal model and provided a number of computational and structural results. Going forward, it would be very interesting to resolve the complexity of Optimal Stable Fractional Matching for the case when agents have *strict* cardinal preferences (i.e., the no-ties case). It would also be very interesting to see if stronger inapproximability results can be obtained for more general matching models, such as stable roommates [59]. Another relevant direction is to consider a stronger solution concept with blocking coalitions of any size (also known as the *core*), wherein the deviating agents can form a fractional matching among themselves to achieve a higher utility for each member of the coalition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Comparison with other notions of stability

In this section, we will discuss other notions of stability of fractional matchings that have been studied for ordinal preferences. For a detailed overview of these and other notions, we refer the reader to the work of Aziz and Klaus [46].

Given an SMC instance $\mathcal{I} = \langle M, W, U, V \rangle$ with cardinal preferences, we can define an ordinal instance $\mathcal{I}^{\text{ord}} = \langle M, W, \succeq \rangle$ where $\succeq := (\succeq_{m_1}, \ldots, \succeq_{m_n}, \succeq_{w_1}, \ldots, \succeq_{w_n})$ is a preference profile consisting of the ordinal preferences of the agents specified as weak total orders. Specifically, for every $i \in \{1, \ldots, n\}$, $\succeq_{m_i}$ and $\succeq_{w_i}$ are weak total orders over the sets W and M, respectively, such that $U(m_i, w) \geq U(m_i, w') \Leftrightarrow w \succeq_{m_i} w'$ and $V(m, w_i) \geq V(m', w_i) \Leftrightarrow m \succeq_{w_i} m'$. We will write $w \succ_m w'$ if $w \succeq_m w'$ but not $w' \succeq_m w$. The relation $m \succ_w m'$ is analogously defined.

Below we will discuss three notions of stability for fractional matchings—*strong stability* [11], *fractional stability* [40], and *ex-post stability* [11]—that were originally formulated in the context of strict ordinal preferences, and were subsequently studied for weak preferences by Aziz and Klaus [46].

The first notion called *strong stability* asserts that no pair of man and woman should positively engage with agents they like less than each other.

Definition 5 (*Strong stability [11]*). A fractional matching μ is *strongly stable* if there are no $m, m' \in M$ and $w, w' \in W$ such that $\mu(m, w') > 0$, $\mu(m', w) > 0$, $w \succ_m w'$, and $m \succ_w m'$.

Next, recall that a fractional matching can be interpreted as a lottery (or a probability distribution) over integral matchings. Under this interpretation, *ex-post stability* requires that every realization of this lottery should be stable.

Definition 6 (*Ex-post stability [11]*). A fractional matching μ is *ex-post stable* if it can be expressed as a convex combination of integral stable matchings.

The third notion, *fractional stability*, formalizes the idea that "if man m is matched with someone less preferred than woman w, then w should be matched with someone more preferred than m" as a linear constraint.

Definition 7 (*Fractional stability [40]*). A fractional matching μ is *fractionally stable* if for every $(m, w) \in M \times W$,

$$\mu(m, w) + \sum_{w' \in W : w' \succeq_m w} \mu(m, w') + \sum_{m' \in M : m' \succeq_w m} \mu(m', w) \geq 1.$$

Notice that an integral matching satisfies the *ordinal* notions of strong (Definition 5), ex-post (Definition 6), or fractional (Definition 7) stability in $\mathcal{I}^{\mathrm{ord}}$ if and only if it satisfies the *cardinal* notion of stability (Definition 1) in $\mathcal{I}$. Thus, the distinction between these notions is meaningful only for fractional matchings. The following observations describe the relationship between these notions (also refer to Fig. 8).

- It is known that strong stability implies ex-post stability even under weak ordinal preferences, but the converse is not true even under strict ordinal preferences [11,46]. Additionally, for strict preferences, ex-post stability is known to be equivalent to fractional stability [40,41,11,22], but for weak preferences, the former is strictly stronger [46].
- A stable matching (Definition 1) may not be fractionally stable (Definition 7) even under strict preferences. To see this, consider the fractional matching μ in Example 1, which was shown to be stable. For the pair (m_1, w_3), we have

$$\mu(m_1, w_3) + \sum_{w' : w' \succ_{m_1} w_3} \mu(m_1, w') + \sum_{m' : m' \succ_{w_3} m_1} \mu(m', w_3) = 0 + 0 + 0.5 < 1,$$

  implying that μ is not fractionally stable (and thus also not ex-post/strongly stable).

Remark 3. It is interesting to note the welfare implication of the above observation. Since fractionally stable matchings are equivalent to ex-post stable matchings under strict preferences, and since all integral stable matchings in Example 1 have social welfare of 7, it follows that the social welfare of any fractionally stable matching is also equal to 7. On the other hand, there exists a stable matching μ with a strictly higher welfare. This illustrates that fractionally stable (and, in particular, strongly stable) matchings can be strictly suboptimal in the cardinal model, further justifying the need to study the computational aspects of optimal stable matchings.

- Given an SMC instance $\mathcal{I}$, a strongly stable (Definition 5) matching in the corresponding ordinal instance $\mathcal{I}^{\mathrm{ord}}$ can be shown to be stable (Definition 1) in the original instance $\mathcal{I}$. Indeed, if μ is strongly stable in $\mathcal{I}^{\mathrm{ord}}$, then for any pair $(m, w) \in M \times W$, we have that either $\sum_{w' : w' \succ_m w'} \mu(m, w') = 0$ or $\sum_{m' : m' \succ_w m'} \mu(m', w) = 0$. Suppose, without loss of generality, that the former holds. Then, the utility of man m in $\mathcal{I}$ is given by

$$u_m(\mu) = U(m, w)\mu(m, w) + \sum_{w' : w' \succeq_m w} U(m, w')\mu(m, w')$$

$$\geq U(m, w) \left(\mu(m, w) + \sum_{w' : w' \succ_m w} \mu(m, w') \right)$$

$$= U(m, w),$$

  as desired.
- As mentioned previously in Remark 1, Narang and Narahari [55] have observed that the set of stable matchings is non-convex even under strict cardinal preferences. Their counterexample involves the convex combination of integral stable matchings, and therefore also establishes that an ex-post stable (and hence also fractionally stable) matching may not be stable.

Appendix B. Non-convexity of strong stability may not imply non-convexity of stability

In this section, we show that non-convexity of strong stability (Definition 5) might not imply the same for stability (Definition 1), even though, as observed in Appendix A, the former is a strictly stronger notion. To this end, we revisit the counterexample used by Aziz and Klaus [46] in establishing the non-convexity of the set of strongly stable matchings. Recall from Appendix A that a fractional matching μ is strongly stable if there are no $m, m' \in M$ and $w, w' \in W$ such that $\mu(m, w') > 0$, $\mu(m', w) > 0$, $w \succ_m w'$, and $m \succ_w m'$.

The counterexample of Aziz and Klaus [46, Theorem 1], which in turn is adapted from [11, Example 2], consists of three men m_1, m_2, m_3 and three women w_1, w_2, w_3 with the following ordinal preferences:

$$m_1 : w_1 \succ w_2 \succ w_3 \qquad w_1 : m_2 \succ m_3 \succ m_1$$
$$m_2 : w_2 \succ w_3 \succ w_1 \qquad w_2 : m_3 \succ m_1 \succ m_2$$
$$m_3 : w_3 \succ w_1 \succ w_2 \qquad w_3 : m_1 \succ m_2 \succ m_3.$$

Consider the integral matchings $\mu^{(1)}$, $\mu^{(2)}$, and $\mu^{(3)}$ defined as follows:

$$\mu^{(1)} := \{(m_1, w_1), (m_2, w_2), (m_3, w_3)\},$$
$$\mu^{(2)} := \{(m_1, w_3), (m_2, w_1), (m_3, w_2)\}, \text{ and}$$
$$\mu^{(3)} := \{(m_1, w_2), (m_2, w_3), (m_3, w_1)\}.$$

Notice that $\mu^{(1)}$, $\mu^{(2)}$, and $\mu^{(3)}$ are strongly stable (and therefore also stable). Also note that the fractional matching $\mu := \frac{1}{3}\mu^{(1)} + \frac{1}{3}\mu^{(2)} + \frac{1}{3}\mu^{(3)}$ violates strong stability for the agents m_1, m_2 and w_2, w_3 since $\mu(m_1, w_3) > 0$, $\mu(m_2, w_2) > 0$, $w_2 \succ_{m_1} w_3$ and $m_1 \succ_{w_2} m_2$. That is, the convex combination of strongly stable integral matchings $\mu^{(1)}$, $\mu^{(2)}$, and $\mu^{(3)}$ is not strongly stable.

Let us now consider an SMC instance $\mathcal{I} = \langle M, W, U, V \rangle$ for the same set of agents, i.e., $M = \{m_1, m_2, m_3\}$, $W = \{w_1, w_2, w_3\}$, and the following valuations:

$$U = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix} \text{ and } V = \begin{bmatrix} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix}.$$

Observe that the valuations in $\mathcal{I}$ are consistent with the aforementioned ordinal preferences. It is easy to see that the utility of each agent in μ is equal to 1, and that μ is stable for $\mathcal{I}$.

Appendix C. Hardness for ε-stability

The proof of Theorem 8 follows along very similar lines to the proof of Theorem 7, again using a reduction from 2P2N-3SAT.

C.1. The reduction

Starting from an instance of 2P2N-3SAT, we first preprocess and augment it in the following way. For each variable of the original instance, we create a copy-variable and, for each clause of the original instance, we create a copy-clause that contains the copy-variables corresponding to the variables of the original clause. Each variable and its copy are *coupled variables* and, similarly, each clause and its copy are *coupled clauses*.

Let the modified input consist of N (boolean) variables $x_1, x_2, \ldots, x_N$, and a 3-CNF formula ϕ with $L = 4N/3$ clauses $c_1, c_2, \ldots, c_{4N/3}$. Note that if ϕ is not satisfiable, then, due to the instance augmentation, there exist at least two clauses that are not satisfied.

We now proceed to describe the instance $\mathcal{I} = \langle M, W, U, V \rangle$ of OPTIMAL ε-STABLE FRACTIONAL MATCHING whose graph representation consists of variable gadgets, clause gadgets, VC-connectors, an accumulator, and CA-connectors. We remark that $\mathcal{I}$ is not an instance with ternary valuations as we use valuations from the set $\{0, 1, \beta, \gamma\}$. We set $\beta = 2(1 - \varepsilon)$, and observe that, when $\varepsilon < 0.03$, we have that $(3\beta^2 + 4)\varepsilon < 1/2$; the value of γ will be set later. Again, we denote by n the number of men (or women) in $\mathcal{I}$; the following reduction is such that $n = \mathcal{O}(N)$.

For each gadget, we classify the edges into the following three types:

- *balanced* edges (m, w) with $U(m, w) = V(m, w) = 1$,
- *man-heavy* edges (m, w) with $U(m, w) > 0$ and $V(m, w) = 0$, and
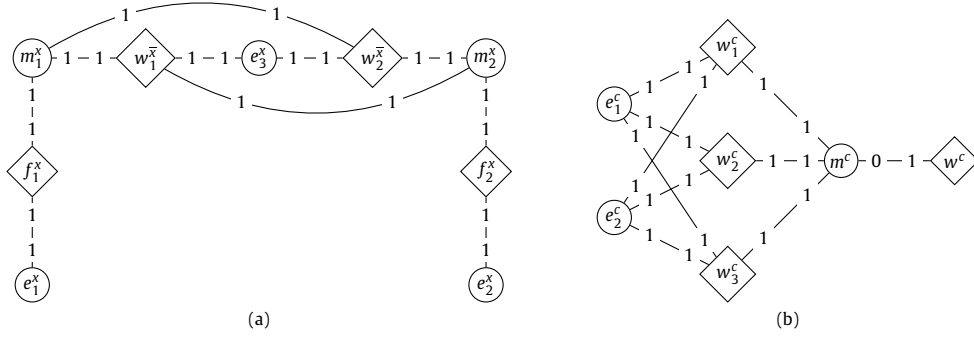- *woman-heavy* edges (m, w) with $U(m, w) = 0$ and $V(m, w) > 0$.

Fig. 9. (a) The variable gadget corresponding to the variable x. (b) The clause gadget corresponding to the clause c and its CA-connector (m^c, w^c).
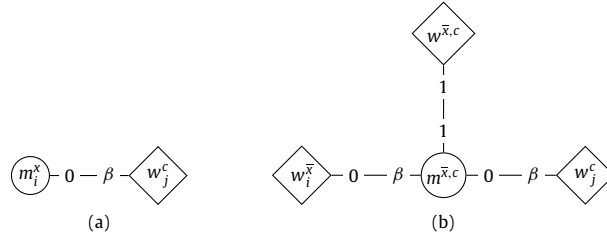

Fig. 10. VC-connectors corresponding to (a) clause c and positive literal x, and (b) clause c and negative literal $\overline{x}$.
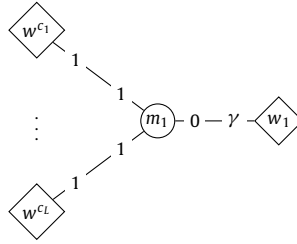

Fig. 11. The accumulator.

Any other pair (m, w) that does not appear as an edge in the graph representation has $U(m, w) = V(m, w) = 0$.

The instance $\mathcal{I}$ contains a variable gadget for every variable x, which consists of five men $m_1^x, m_2^x, e_1^x, e_2^x, e_3^x$, four women $w_1^{\overline{x}}, w_2^{\overline{x}}, f_1^x, f_2^x$ and the ten balanced edges $(e_1^x, f_1^x), (m_1^x, f_1^x), (m_1^x, w_1^{\overline{x}}), (e_3^x, w_1^{\overline{x}}), (e_3^x, w_2^{\overline{x}}), (m_2^x, w_2^{\overline{x}}), (m_2^x, f_2^x), (e_2^x, f_2^x), (m_1^x, w_2^{\overline{x}})$, and $(m_2^x, w_1^{\overline{x}})$ (see Fig. 9a).

For every clause c, instance $\mathcal{I}$ has a clause gadget with three men m^c, e_1^c, e_2^c, three women w_1^c, w_2^c, w_3^c, and the nine balanced edges between them (see Fig. 9b).

For every appearance of a literal in a clause, there is a variable-clause connector (or VC-connector) whose structure depends on whether it corresponds to a positive or a negative literal. In particular, for every positive literal x whose i-th appearance ($i \in \{1, 2\}$) is as the j-th literal ($j \in \{1, 2, 3\}$) of clause c, instance $\mathcal{I}$ has a VC-connector that consists of a single woman-heavy edge between m_i^x (from the variable gadget corresponding to variable x) and w_j^c (from the clause gadget corresponding to clause c) such that $U(m_i^x, w_j^c) = 0$ and $V(m_i^x, w_j^c) = \beta$ (see Fig. 10a). This edge is simultaneously the input and output edge of the VC-connector.

Similarly, for every negative literal $\overline{x}$ whose i-th appearance ($i \in \{1, 2\}$) is as the j-th literal ($j \in \{1, 2, 3\}$) of clause c, instance $\mathcal{I}$ has a VC-connector that consists of man $m^{\overline{x}, c}$, woman $w^{\overline{x}, c}$, the man-heavy edge $(m^{\overline{x}, c}, w_i^{\overline{x}})$ with $U(m^{\overline{x}, c}, w_i^{\overline{x}}) = \beta$ and $V(m^{\overline{x}, c}, w_i^{\overline{x}}) = 0$, the balanced edge $(m^{\overline{x}, c}, w^{\overline{x}, c})$, and the woman-heavy edge $(m^{\overline{x}, c}, w_j^c)$ with $U(m^{\overline{x}, c}, w_j^c) = 0$ and $V(m^{\overline{x}, c}, w_j^c) = \beta$ (see Fig. 10b). Here, $(m^{\overline{x}, c}, w_i^{\overline{x}})$ is the input and $(m^{\overline{x}, c}, w_j^c)$ is the output edge.

The accumulator gadget consists of man m_1, woman w_1 and woman w^c for every clause c of ϕ. In addition, it contains the balanced edges (m_1, w^c) for every clause c, which we once again call *tine* edges and the woman-heavy edge (m_1, w_1) with $U(m_1, w_1) = 0$ and $V(m_1, w_1) = \gamma$ (see Fig. 11).

Finally, instance $\mathcal{I}$ contains a clause-accumulator connector (or CA-connector) for every clause c consisting of the woman-heavy edge (m^c, w^c) between the man m^c (from the clause gadget corresponding to clause c) and woman w^c (from the accumulator) so that $U(m^c, w^c) = 0$ and $V(m^c, w^c) = 1$ (see Fig. 9b). Any other edge (m, w) is such that $U(m, w) = V(m, w) = 0$.

Notice that the above construction has more women than men. To restore balance, we pad the instance with extra (isolated) men that neither value nor are valued by any other agent. This completes the construction of the reduced instance.

C.2. Gadget properties

We will now prove several important properties (Claims 5-8) of our construction.

Claim 5. *For every variable x, an ε-stable fractional matching μ satisfies at least one of the following conditions:*

(1) $\mu(m_1^x, w_1^{\overline{x}}) + \mu(m_1^x, w_2^{\overline{x}}) + \mu(m_1^x, f_1^x) \geq 1 - \varepsilon$ *and* $\mu(m_2^x, w_1^{\overline{x}}) + \mu(m_2^x, w_2^{\overline{x}}) + \mu(m_2^x, f_2^x) \geq 1 - \varepsilon$.
(2) $\mu(m_1^x, w_1^{\overline{x}}) + \mu(m_2^x, w_1^{\overline{x}}) + \mu(e_3^x, w_1^{\overline{x}}) \geq 1 - \varepsilon$ *and* $\mu(m_1^x, w_2^{\overline{x}}) + \mu(m_2^x, w_2^{\overline{x}}) + \mu(e_3^x, w_2^{\overline{x}}) \geq 1 - \varepsilon$.

Proof. Assume otherwise that, for some $i, j \in \{1, 2\}$, $\mu(m_i^x, w_1^{\overline{x}}) + \mu(m_i^x, w_2^{\overline{x}}) + \mu(m_i^x, f_i^x) < 1 - \varepsilon$ and $\mu(m_1^x, w_j^{\overline{x}}) + \mu(m_2^x, w_j^{\overline{x}}) + \mu(e_3^x, w_j^{\overline{x}}) < 1 - \varepsilon$. Then, since instance $\mathcal{I}$ contains no man-heavy edge (m_i^x, w) and no woman-heavy edge $(m, w_j^{\overline{x}})$, both m_i^x and $w_j^{\overline{x}}$ have utility less than $1 - \varepsilon$ in μ and hence the pair $(m_i^x, w_j^{\overline{x}})$ is ε-blocking—a contradiction. □

As before, the two conditions in the statement of Claim 5 affect the weight of the input edges of the VC-connectors that are attached to the variable gadget in any ε-stable fractional matching. Thus, condition (1) implies that the weight assigned to each input edge of the VC-connectors that correspond to the two appearances of the positive literal x in clauses must be at most ε. To see why, observe that these input edges are incident to nodes m_1^x and m_2^x, and the total weight of all edges incident to each of these nodes cannot exceed 1. Similarly, condition (2) implies that the weight assigned to each input edge of the VC-connectors that correspond to the two appearances of the negative literal $\overline{x}$ in clauses must be at most ε.

Claim 6. *Any ε-stable fractional matching that assigns a weight of at most ε to the input edge of a VC-connector must assign a weight of at most $\beta\varepsilon$ to its output edge as well.*

Proof. The claim holds trivially for VC-connectors corresponding to positive literals. Consider a VC-connector corresponding to a negative literal $\overline{x}$ and a clause c containing it, and let $\zeta \leq \varepsilon$ be the weight assigned to the input edge $(m^{\overline{x},c}, w^{\overline{x}})$. Observe that, besides the input edge $(m^{\overline{x},c}, w^{\overline{x}})$, the edge $(m^{\overline{x},c}, w^{\overline{x},c})$ is the only balanced or man-heavy edge that is incident to man $m^{\overline{x},c}$ and the only balanced (or woman-heavy) edge incident to woman $w^{\overline{x},c}$. Hence, ε-stability of edge $(m^{\overline{x},c}, w^{\overline{x},c})$ requires a weight of at least $1 - \varepsilon - \beta\zeta$ assigned to it. Then, the output edge, which is also incident to node $m^{\overline{x},c}$, must have a weight of at most $\varepsilon + (\beta - 1)\zeta \leq \beta\varepsilon$. □

Claim 7. *Any ε-stable fractional matching that assigns a weight of at most $\beta\varepsilon$ to each output edge of the VC-connectors corresponding to clause c must assign a weight of at most $3(\beta^2 + 1)\varepsilon$ to the CA-connector of clause c as well.*

Proof. Let ℓ_1, ℓ_2, and ℓ_3 be the literals of clause c. Consider, for the sake of contradiction, an ε-stable fractional matching μ that assigns (1) a weight of at most $\beta\varepsilon$ to each output edge of the VC-connectors corresponding to literals ℓ_i and clause c, and (2) a weight of more than $3(\beta^2 + 1)\varepsilon$ to edge (m^c, w^c) of the CA-connector for clause c. Note that condition (2) implies that the total weight on the edges (m^c, w_1^c), (m^c, w_2^c) and (m^c, w_3^c) is strictly smaller than $1 - 3(\beta^2 + 1)\varepsilon$. Since these are the only balanced or man-heavy edges incident to man m^c, the ε-stability of these edges is guaranteed by a utility of (at least) $1 - \varepsilon$ for each of the agents w_1^c, w_2^c, and w_3^c.

Note that besides the output edges of the VC-connectors, the edges (e_1^c, w_i^c), (e_2^c, w_i^c), and (m^c, w_i^c) are the only balanced or woman-heavy edges incident to agent w_i^c for all $i \in \{1, 2, 3\}$. Along with condition (1), this implies that the weight assigned to these three edges is at least $1 - (\beta^2 + 1)\varepsilon$. Hence, the total weight on the nine edges of the clause gadget is at least $3 - 3(\beta^2 + 1)\varepsilon$, which, by the definition of β, is strictly more than 2 for the six edges incident to men e_1^c and e_2^c, violating the definition of a fractional matching. □

Claim 8. *Any ε-stable fractional matching that assigns a weight of at most $3(\beta^2 + 1)\varepsilon$ to at least two CA-connectors must assign a total weight of $1 - \varepsilon$ to the tine edges and a weight of at most ε to the edge (m_1, w_1) of the accumulator.*

Proof. Assume that a weight of at most $3(\beta^2 + 1)\varepsilon$ has been assigned to the edges (m^{c_1}, w^{c_1}) and (m^{c_2}, w^{c_2}) of the CA-connectors corresponding to some clauses c_1 and c_2. Since these are the only edges for which agents w^{c_1} and w^{c_2} have positive value, and there is no man-heavy edge incident to agent m_1, stability on the edges (m_1, w^{c_1}) and (m_1, w^{c_2}) requires that the total weight of the tine edges (m_1, w^c) (for every clause c) is (at least) $1 - \varepsilon$. Indeed, since by the definition of β we have $(3\beta^2 + 4)\varepsilon < 1/2$, it is not possible to guarantee stability of the edges (m_1, w^{c_1}) and (m_1, w^{c_2}) by assigning weight at least $1 - (3\beta^2 + 4)\varepsilon$ to each of them so that both w^{c_1} and w^{c_2} have utility at least $1 - \varepsilon$. Hence, the weight of the edge (m_1, w_1) of the accumulator is at most ε. □

C.3. Proof of inapproximability

Lemma 6. *If formula ϕ is not satisfiable, then any ε-stable fractional matching of $\mathcal{I}$ has welfare at most $56\beta N + \gamma\varepsilon$.*

Proof. We will first show that if ϕ is not satisfiable, then any ε-stable fractional matching of $\mathcal{I}$ assigns weight at most $3(\beta^2+1)\varepsilon$ to at least two CA-connectors. For the sake of contradiction, consider an ε-stable fractional matching that assigns weight at most $3(\beta^2+1)\varepsilon$ to at most one CA-connector; let c be the relevant clause, if such a clause exists. We will construct a truth assignment for formula ϕ (contradicting the assumption of the lemma) by repeating the following process for every clause $c' \neq c$ of ϕ. Let ℓ be a literal that appears in c' such that the output edge of the VC-connector, that corresponds to the appearance of ℓ in c', has weight greater than $\beta\varepsilon$. Recall that such a literal exists by Claim 7. We set ℓ to 1 (true). For every variable that has not received a value in this way, we arbitrarily set it to 1.

The above assignment satisfies all clauses except possibly for clause c. Since clause c is coupled with another clause $\hat{c}$ that is satisfied, it suffices to assign one of the variables appearing in c the same value as its corresponding coupled variable appearing in $\hat{c}$. To show that the assignment is also valid, we need to argue that there is no variable x such that both literals x and $\bar{x}$ have been set to 1. Assume otherwise that this is the case. Furthermore, assume that literal x was set to 1 due to its appearance in a clause c_1, and that this is its i_1-th appearance (with $i_1 \in \{1, 2\}$). Also, literal $\bar{x}$ was set to 1 due to its appearance in a different clause c_2, where $\bar{x}$ makes its i_2-th appearance (again, $i_2 \in \{1, 2\}$). Hence, the output edge of the VC-connector that corresponds to literal x and clause c_1 (respectively, the VC-connector that corresponds to literal $\bar{x}$ and clause c_2) has weight greater than $\beta\varepsilon$. Then, by Claim 6, the input edges of both VC-connectors have weight greater than ε. As these input edges are incident to nodes $m_{i_1}^x$ and $w_{i_2}^{\bar{x}}$, Claim 5 yields that the total weight in the edges incident to some of the nodes $m_{i_1}^x$ and $w_{i_2}^{\bar{x}}$ is strictly higher than 1, contradicting the definition of a fractional matching.

Since any ε-stable fractional matching assigns a weight of at most $3(\beta^2+1)\varepsilon$ to at least two CA-connectors, by Claim 8, the contribution of the accumulator to the welfare is at most $2(1-\varepsilon) + \gamma\varepsilon$ ($2(1-\varepsilon)$ from the tine edges plus $\gamma\varepsilon$ from the accumulator). The upper bound follows by considering the sum of valuations of all agents for edges that do not belong to the accumulator. This sum consists of

- total value of 20 for the ten balanced edges of each of the N variable gadgets,
- total value of β for the edges of each of the $2N$ VC-connectors corresponding to a positive literal,
- total value of $2 + 2\beta$ for the edges of each of the $2N$ VC-connectors corresponding to a negative literal,
- total value of 19 for the nine balanced edges of each of the $4N/3$ clause gadgets and their corresponding CA-connectors.

It can be easily seen that $56\beta N - 2(1-\varepsilon)$ is a (loose) upper bound on the total value from these edges. $\square$

Lemma 7. *If ϕ is satisfiable, then there exists an ε-stable fractional matching on $\mathcal{I}$ that has welfare at least γ.*

Proof. Consider an assignment of boolean values to the variables that satisfies ϕ. We construct an ε-stable fractional matching μ in $\mathcal{I}$ so that the contribution of the accumulator gadget to the welfare is at least γ.

Variable gadgets. The weights on the edges of the variable gadget corresponding to variable x are:

- $\mu(m_1^x, w_1^{\bar{x}}) = \mu(e_3^x, w_1^{\bar{x}}) = \mu(e_3^x, w_2^{\bar{x}}) = \mu(m_2^x, w_2^{\bar{x}}) = 1/2$, $\mu(e_1^x, f_1^x) = \mu(e_2^x, f_2^x) = 1$, and $\mu(m_1^x, f_1^x) = \mu(m_1^x, w_2^{\bar{x}}) = \mu(m_2^x, f_2^x) = \mu(m_2^x, w_1^{\bar{x}}) = 0$ if x is true, and
- $\mu(e_3^x, w_1^{\bar{x}}) = \mu(e_3^x, w_2^{\bar{x}}) = 1/2$, $\mu(m_1^x, f_1^x) = \mu(m_2^x, f_2^x) = 1$, $\mu(m_1^x, w_1^{\bar{x}}) = \mu(m_2^x, w_2^{\bar{x}}) = \mu(m_1^x, w_2^{\bar{x}}) = \mu(m_2^x, w_1^{\bar{x}}) = \mu(e_1^x, f_1^x) = \mu(e_2^x, f_2^x) = 0$ if x is false.

Clause gadgets and CA-connectors. For every clause we select arbitrarily one of the true literals of the clause and call it *active*; since the assignment satisfies ϕ, there is certainly such a literal. Consider clause c and let ℓ_i (with $i \in \{1, 2, 3\}$) be its active literal; let i_1 and i_2 be the indices from $\{1, 2, 3\}$ than are different than i. We set $\mu(e_1^c, w_{i_1}^c) = \mu(e_2^c, w_{i_2}^c) = 1$ and $\mu(e_1^c, w_i^c) = \mu(e_1^c, w_{i_2}^c) = \mu(e_2^c, w_i^c) = \mu(e_2^c, w_{i_1}^c) = \mu(m^c, w_i^c) = \mu(m^c, w_{i_1}^c) = \mu(m^c, w_{i_2}^c) = 0$. We also assign a weight of 1 to the CA-connector corresponding to c, i.e., $\mu(m^c, w^c) = 1$.

VC-connectors. For every non-active VC-connector, we set the weight of its balanced edge, if it exists, to 1 and the weight of the remaining edges to 0. For every active VC-connector corresponding to the i-th appearance of the positive literal x as the j-th literal of clause c ($i \in \{1, 2\}$, $j \in \{1, 2, 3\}$), we set $\mu(m_i^x, w_j^c) = 1/2$.

For every active VC-connector corresponding to the i-th appearance of the negative literal x as the j-th literal of clause c ($i \in \{1, 2\}$, $j \in \{1, 2, 3\}$), we set the weights of its edges as follows: $\mu(m^{\bar{x},c}, w_i^{\bar{x}}) = \mu(m^{\bar{x},c}, w_j^c) = 1/2$ and $\mu(m^{\bar{x},c}, w^{\bar{x},c}) = 0$.

Accumulator. We set $\mu(m_1, w^c) = 0$ for every tine edge (m_1, w^c) of the accumulator. Furthermore, we set $\mu(m_1, w_1) = 1$. So, the contribution of the accumulator to the social welfare is γ, as desired.

It can be easily verified that the total weight of the edges that are incident to any node is at most 1. Hence, μ is a valid fractional matching. Regarding stability, it suffices to verify that either the man or the woman of a balanced pair has

a utility of at least $1 - \varepsilon$. Note that we only need to check for the balanced edges, as the man- or woman-heavy edges and the remaining pairs do not impose any constraints on stability. For the balanced edges, stability is established by the following series of observations (we will use the term 'stabilized by' to denote that an agent's utility is at least $1 - \varepsilon$): The variable gadget for the variable x (Fig. 9a) is stabilized by the agents f_1^x, f_2^x, e_3^x along with m_1^x, m_2^x (if x is true) or $w_1^{\bar{x}}$, $w_2^{\bar{x}}$ (if x is false). The clause gadget for clause c (Fig. 9b) with active index i (and non-active indices i_1 and i_2) is stabilized by the agents e_1^c, e_2^c, w_i^c, $w_{i_1}^c$, $w_{i_2}^c$; in particular, the edge (m^c, w_i^c) is stabilized by w_i^c because an active literal triggers the woman-heavy edge in the VC-connector. A VC connector is stabilized by $m^{\bar{x},c}$ (Fig. 10b). Finally, the tine edges in the accumulator (Fig. 11) are stabilized by $w^{c_1}, \ldots, w^{c_L}$ (because we trigger the CA-connector). Overall, μ is a feasible stable fractional matching. □

We are ready to prove Theorem 8. We select a value of γ such that $\gamma \geq \frac{56\beta N(1/\varepsilon-\delta)}{\varepsilon\delta}$. By Lemma 6, the welfare of μ if ϕ was not satisfiable would be at most

$$56\beta N + \gamma\varepsilon \leq \frac{\gamma\varepsilon\delta}{1/\varepsilon - \delta} + \gamma\varepsilon = \frac{\gamma}{1/\varepsilon - \delta}.$$

By Lemma 7, we have that the welfare of μ if ϕ was not satisfiable would be at least $1/\varepsilon - \delta$ times smaller than the welfare $\mathcal{I}$ could have if ϕ was satisfiable.

Appendix D. Approximation algorithm for $\frac{1}{2}$-stability

We will now provide a polynomial-time algorithm that computes a $\frac{1}{2}$-stable fractional matching with welfare at least that of an optimal (exactly) stable fractional matching. Notice that unlike Theorems 5 and 6, where the quality of the computed matching is compared to the optimal matching μ^{opt}, the guarantee in Theorem 9 is considerably weaker.

Theorem 9. *Let $\mathcal{I}$ be an SMC instance and μ^* be an optimal stable fractional matching for $\mathcal{I}$. Then, a $\frac{1}{2}$-stable fractional matching μ that satisfies $\mathcal{W}(\mu) \geq \mathcal{W}(\mu^*)$ can be computed in polynomial time.*

Proof. Consider the mixed integer linear program (OPT-Stab) from Section 2.1 for finding an optimal stable fractional matching for $\mathcal{I}$. Relaxing the integrality constraint (8) to $y(m, w) \in [0, 1]$ results in a linear program. Since a stable fractional matching always exists (see Proposition 1), this relaxation is feasible. Let μ be a solution of the relaxed program. Since $\max\{y(m, w), 1 - y(m, w)\} \geq \frac{1}{2}$, we have that for every man-woman pair $(m, w) \in M \times W$, either $u_m \geq \frac{1}{2}U(m, w)$ or $v_w \geq \frac{1}{2}V(m, w)$, implying that μ is $\frac{1}{2}$-stable. It is also clear that $\mathcal{W}(\mu) \geq \mathcal{W}(\mu^*)$ since μ^* satisfies (OPT-Stab). □

References

[1] D. Gale, L.S. Shapley, College admissions and the stability of marriage, Am. Math. Mon. 69 (1) (1962) 9–15.
[2] D. Gusfield, R.W. Irving, The Stable Marriage Problem: Structure and Algorithms, MIT Press, 1989.
[3] A.E. Roth, M.A.O. Sotomayor, Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis, Cambridge University Press, 1990.
[4] D.E. Knuth, Stable Marriage and Its Relation to Other Combinatorial Problems: An Introduction to the Mathematical Analysis of Algorithms, Vol. 10, American Mathematical Soc., 1997.
[5] D.F. Manlove, Algorithmics of Matching Under Preferences, Vol. 2, World Scientific, 2013.
[6] A.E. Roth, The evolution of the labor market for medical interns and residents: a case study in game theory, J. Polit. Econ. 92 (6) (1984) 991–1016.
[7] A.E. Roth, E. Peranson, The redesign of the matching market for American physicians: some engineering aspects of economic design, Am. Econ. Rev. 89 (4) (1999) 748–780.
[8] A. Abdulkadiroğlu, P.A. Pathak, A.E. Roth, The New York City High School Match, Am. Econ. Rev. (2005) 364–367.
[9] A.E. Roth, T. Sönmez, M.U. Ünver, Kidney exchange, Q. J. Econ. 119 (2) (2004) 457–488.
[10] X. Wang, N. Agatz, A. Erera, Stable matching for dynamic ride-sharing systems, Transp. Sci. 52 (4) (2018) 850–867.
[11] A.E. Roth, U.G. Rothblum, J.H. Vande Vate, Stable matchings, optimal assignments, and linear programming, Math. Oper. Res. 18 (4) (1993) 803–828.
[12] A. Bogomolnaia, H. Moulin, Random matching under dichotomous preferences, Econometrica 72 (1) (2004) 257–279.
[13] S. Bronfman, N. Alon, A. Hassidim, A. Romm, Redesigning the Israeli Medical Internship Match, ACM Trans. Econ. Comput. 6 (3–4) (2018) 1–18.
[14] J. Gudmundsson, Compromises and rewards: stable and non-manipulable probabilistic matching, Int. J. Game Theory 48 (2) (2019) 365–392.
[15] Z. Li, K. Lieberman, W. Macke, S. Carrillo, C.-J. Ho, J. Wellen, S. Das, Incorporating compatible pairs in kidney exchange: a dynamic weighted matching model, in: Proceedings of the 2019 ACM Conference on Economics and Computation, ACM, 2019, pp. 349–367.
[16] E. Anshelevich, S. Das Matching, Cardinal utility, and social welfare, ACM SIGecom Exch. 9 (1) (2010) 4.
[17] M.S. Pini, F. Rossi, K.B. Venable, T. Walsh, Stability, optimality and manipulation in matching problems with weighted preferences, Algorithms 6 (4) (2013) 782–804.
[18] F. Echenique, A.J. Wilson, L. Yariv, Clearinghouses for two-sided matching: an experimental study, Quant. Econ. 7 (2) (2016) 449–482.
[19] E. Anshelevich, S. Das, Y. Naamad, Anarchy, stability, and utopia: creating better matchings, Auton. Agents Multi-Agent Syst. 26 (1) (2013) 120–140.
[20] D. Gusfield, Three fast algorithms for four problems in stable marriage, SIAM J. Comput. 16 (1) (1987) 111–128.
[21] R.W. Irving, P. Leather, D. Gusfield, An efficient algorithm for the "optimal" stable marriage, J. ACM 34 (3) (1987) 532–543.
[22] C.-P. Teo, J. Sethuraman, The geometry of fractional stable matchings and its applications, Math. Oper. Res. 23 (4) (1998) 874–891.
[23] D.F. Manlove, R.W. Irving, K. Iwama, S. Miyazaki, Y. Morita, Hard variants of stable marriage, Theor. Comput. Sci. 276 (1–2) (2002) 261–279.
[24] T. Mai, V.V. Vazirani, A natural generalization of stable matching solved via new insights into ideal cuts, preprint, arXiv:1802.06621.
[25] A. Kato, Complexity of the sex-equal stable marriage problem, Jpn. J. Ind. Appl. Math. 10 (1) (1993) 1.
[26] T. Feder, A new fixed point approach for stable networks and stable marriages, J. Comput. Syst. Sci. 45 (2) (1992) 233–284.

[27] K. Iwama, S. Miyazaki, Y. Morita, D.F. Manlove, Stable marriage with incomplete lists and ties, in: International Colloquium on Automata, Languages, and Programming, Springer, 1999, pp. 443–452.
[28] R.W. Irving, D.F. Manlove, G. O'Malley, Stable marriage with ties and bounded length preference lists, J. Discret. Algorithms 7 (2) (2009) 213–219.
[29] J. Drummond, C. Boutilier, Elicitation and approximately stable matching with partial preferences, in: Twenty-Third International Joint Conference on Artificial Intelligence, 2013, pp. 97–105.
[30] B. Rastegari, P. Goldberg, D.F. Manlove, Preference elicitation in matching markets via interviews: a study of offline benchmarks, in: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, 2016, pp. 1393–1394.
[31] M. Goto, A. Iwasaki, Y. Kawasaki, R. Kurata, Y. Yasuda, M. Yokoo, Strategyproof matching with regional minimum and maximum quotas, Artif. Intell. 235 (2016) 40–57.
[32] N. Hamada, C.-L. Hsu, R. Kurata, T. Suzuki, S. Ueda, M. Yokoo, Strategy-proof school choice mechanisms with minimum quotas and initial endowments, Artif. Intell. 249 (2017) 47–71.
[33] I.P. Gent, P. Prosser, B. Smith, T. Walsh, SAT encodings of the stable marriage problem with ties and incomplete lists, in: Fifth International Symposium on the Theory and Applications of Satisfiability Testing, 2002, pp. 133–140.
[34] J. Drummond, A. Perrault, F. Bacchus, SAT is an effective and complete method for solving stable matching problems with couples, in: Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015, pp. 518–525.
[35] A. Perrault, J. Drummond, F. Bacchus, Strategy-proofness in the stable matching problem with couples, in: Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems, 2016, pp. 132–140.
[36] I.P. Gent, R.W. Irving, D.F. Manlove, P. Prosser, B.M. Smith, A constraint programming approach to the stable marriage problem, in: International Conference on Principles and Practice of Constraint Programming, Springer, 2001, pp. 225–239.
[37] D.F. Manlove, G. O'Malley, P. Prosser, C. Unsworth, A constraint programming approach to the hospitals/residents problem, in: International Conference on Integration of Artificial Intelligence (AI) and Operations Research (OR) Techniques in Constraint Programming, Springer, 2007, pp. 155–170.
[38] D.F. Manlove, I. McBride, J. Trimble, "Almost-stable" matchings in the hospitals/residents problem with couples, Constraints 22 (1) (2017) 50–72.
[39] U. Endriss, Analysis of one-to-one matching mechanisms via SAT solving: impossibilities for universal axioms, in: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI Press, 2020, pp. 1918–1925.
[40] J.H. Vande Vate, Linear programming brings marital bliss, Oper. Res. Lett. 8 (3) (1989) 147–153.
[41] U.G. Rothblum, Characterization of stable matchings as extreme points of a polytope, Math. Program. 54 (1–3) (1992) 57–67.
[42] R.V. Vohra, Stable matchings and linear programming, Curr. Sci. 105 (9) (2012) 1051–1055.
[43] X. Chen, G. Ding, X. Hu, W. Zang, The maximum-weight stable matching problem: duality and efficiency, SIAM J. Discrete Math. 26 (3) (2012) 1346–1360.
[44] R. Aharoni, T. Fleiner, On a lemma of Scarf, J. Comb. Theory, Ser. B 87 (1) (2003) 72–80.
[45] P. Biró, T. Fleiner, Fractional solutions for capacitated NTU-games, with applications to stable matchings, Discrete Optim. 22 (2016) 241–254.
[46] H. Aziz, B. Klaus, Random matching under priorities: stability and no envy concepts, Soc. Choice Welf. (2019) 1–47.
[47] A. Deligkas, G.B. Mertzios, P.G. Spirakis, The computational complexity of weighted greedy matching, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI Press, 2017, pp. 317–325.
[48] V. Manjunath, Stability and the Core of Probabilistic Marriage Problems, Available at SSRN 1809941.
[49] J.W. Hatfield, S.D. Kominers, A. Nichifor, M. Ostrovsky, A. Westkamp, Stability and competitive equilibrium in trading networks, J. Polit. Econ. 121 (5) (2013) 966–1005.
[50] J.W. Hatfield, S.D. Kominers, Multilateral matching, J. Econ. Theory 156 (2015) 175–206.
[51] B. Doğan, K. Yıldız, Efficiency and stability of probabilistic assignments in marriage problems, Games Econ. Behav. 95 (2016) 47–58.
[52] F. Echenique, A. Galichon, Ordinal and cardinal solution concepts for two-sided matching, Games Econ. Behav. 101 (2017) 63–77.
[53] K.C. Ágoston, P. Biró, I. McBride, Integer programming methods for special college admissions problems, J. Comb. Optim. 32 (4) (2016) 1371–1399.
[54] M. Delorme, S. García, J. Gondzio, J. Kalcsics, D. Manlove, W. Pettersson, Mathematical models for stable matching problems with ties and incomplete lists, Eur. J. Oper. Res. 277 (2) (2019) 426–441.
[55] S. Narang, Y. Narahari, A study of incentive compatibility and stability issues in fractional matchings, preprint, arXiv:2001.05652.
[56] R.J. Lipton, E. Markakis, A. Mehta, Playing large games using simple strategies, in: Proceedings of the Fourth ACM Conference on Electronic Commerce, ACM, 2003, pp. 36–41.
[57] S. Barman, Approximating Nash equilibria and dense subgraphs via an approximate version of Carathéodory's theorem, SIAM J. Comput. 47 (3) (2018) 960–981.
[58] R. Yoshinaka, Higher-order matching in the linear lambda calculus in the absence of constants is NP-complete, in: Proceedings of the Sixteenth International Conference on Term Rewriting and Applications, Springer-Verlag, 2005, pp. 235–249.
[59] R.W. Irving, An efficient algorithm for the "stable roommates" problem, J. Algorithms 6 (4) (1985) 577–595.