Balanced Stable Marriage: How Close is Close Enough?*

Sushmita Gupta†   Sanjukta Roy‡   Saket Saurabh†‡   Meirav Zehavi†

Abstract

The BALANCED STABLE MARRIAGE problem is a central optimization version of the classic STABLE MARRIAGE problem. Here, the output cannot be an arbitrary stable matching, but one that balances between the dissatisfaction of the two parties, men and women. We study BALANCED STABLE MARRIAGE from the viewpoint of Parameterized Complexity. Our "above guarantee parameterizations" are arguably the most natural parameterizations of the problem at hand. Indeed, our parameterizations precisely fit the scenario where there exists a stable marriage that both parties would accept, that is, where the satisfaction of each party is *close* to the best it can hope for. Furthermore, our parameterizations accurately draw the line between tractability and intractability with respect to the target value.

1 Introduction

Matching under preferences is a rich topic central to both economics and computer science, which has been consistently and intensively studied for over several decades. One of the main reasons for interest in this topic stems from the observation that it is extremely relevant to a wide variety of practical applications modeling situations where the objective is to *match* agents to other agents (or to resources). In the most general setting, a matching is defined as an allocation (or assignment) of agents to resources that satisfies some predefined criterion of compatibility/acceptability. Here, the (arguably) best known model is the *two-sided model*, where the agents on one side are referred to as *men*, and the agents on the other side are referred to as *women*. A few illustrative examples of real life situations where this model is employed in practice include matching hospitals to residents, students to colleges, kidney patients to donors and users to servers in a distributed Internet service. At the heart of all of these applications lies the fundamental STABLE MARRIAGE problem. In particular, the Nobel Prize in Economics was awarded to Shapley and Roth in 2012 "for the theory of stable allocations and the practice of market design." Moreover, several books have been dedicated to the study of STABLE MARRIAGE as well as optimization variants of this classical problem such as the EGALITARIAN STABLE MARRIAGE, SEX-EQUAL STABLE MARRIAGE and BALANCED STABLE MARRIAGE problems [11, 14, 16].

The input of the STABLE MARRIAGE problem consists of a set of men, M, and a set of women, W, each person ranking a subset of people of the opposite gender. That is, each person a has a set of *acceptable partners*, $\mathcal{A}(a)$, whom this person subjectively ranks. Consequently, each person a has a so-called *preference list*, where $p_a(b)$ denotes the position of $b \in \mathcal{A}(a)$ in a's preference list. Without loss of generality, it is assumed that if a person a ranks a person b, then the person b ranks the person a as well. The sets of preference lists of the men and the women are denoted by $\mathcal{L}_M$ and $\mathcal{L}_W$, respectively. In this context, we say that a pair of a man and a woman, (m, a), is an *acceptable pair* if both $m \in \mathcal{A}(w)$ and $w \in \mathcal{A}(m)$. Accordingly, the notion of a *matching* refers to a matching between men and women, where two people that are matched

to one another form an acceptable pair. Roughly speaking, the goal of the STABLE MARRIAGE problem is to *find* a matching that is *stable* in the following sense: there should not exist two people who prefer being matched to one another over their current "status". More precisely, a matching μ is said to be stable if it does not have a *blocking pair*, which is an acceptable pair (m, w) such that **(i)** either m is unmatched by μ or $p_m(w) < p_m(\mu(m))$, and **(ii)** either w is unmatched by μ or $p_w(m) < p_w(\mu(w))$. Here, the notation $\mu(a)$ indicates the person to whom μ matches the person a. Note that a person always prefers being matched to an acceptable partner over being unmatched.

The seminal paper [7] by Gale and Shapely on stable matchings shows that given an instance of STABLE MARRIAGE, a stable matching necessarily *exists*, but it is not necessarily unique. In fact, for a given instance of STABLE MARRIAGE, there can be an *exponential* number of stable matchings, and they should be viewed as a *spectrum* where the two extremes are known as the *man-optimal stable matching* and the *woman-optimal stable matching*. Formally, the man-optimal stable matching, denoted by μ_M, is a stable matching such that every stable matching μ satisfies the following condition: every man m is either unmatched by both μ_M and μ or $p_m(\mu_M(m)) \leq p_m(\mu(m))$. The woman-optimal stable matching, denoted by μ_W, is defined analogously. These two extremes, which give the best possible solution for one party at the expense of the other party, always exist and can be computed in polynomial time [7]. Naturally, it is desirable to analyze matchings that lie somewhere in the middle, being *globally desirable*, *fair towards both sides* or *desirable by both sides*.

Each notion above of what constitutes a desirable stable matching leads to a natural, *different* optimization problem. The determination of which notion best describes an appropriate outcome depends on the specific situation at hand. Here, the quantity $p_a(\mu(a))$ is viewed as the "satisfaction" of a in a matching μ, where a smaller value signifies a greater amount of satisfaction. Under this interpretation, the *egalitarian stable matching* attempts to be *globally desirable* by minimizing $e(\mu) = \sum_{(m,w)\in\mu}(p_m(\mu(m)) + p_w(\mu(w)))$ over the set of all stable matchings, which we denote by SM. The problem of finding an egalitarian stable matching, called EGALITARIAN STABLE MARRIAGE, is known be solvable in polynomial time due to Irving et al. [12]. Roughly speaking, this problem does not distinguish between men and women, and therefore it does not fit scenarios where it is necessary to differentiate between the individual satisfaction of each party. In such scenarios, the SEX-EQUAL STABLE MARRIAGE and BALANCED STABLE MARRIAGE problems come into play. Before we define each of these two problems, we would like to remark that a survey of results related to EGALITARIAN STABLE MARRIAGE and SEX-EQUAL STABLE MARRIAGE is outside the scope of this paper, and we refer interested readers to the books [11, 14, 16]. Here, we consider these two problems only to understand the context of BALANCED STABLE MARRIAGE.

In the SEX-EQUAL STABLE MARRIAGE problem, the objective is to find a stable matching that minimizes the absolute value of $\delta(\mu)$ over SM, where $\delta(\mu) = \sum_{(m,w)\in\mu} p_m(\mu(m))$ $- \sum_{(m,w)\in\mu} p_w(\mu(w))$. It is thus clear that SEX-EQUAL STABLE MARRIAGE seeks a stable matching that is fair towards both sides by minimizing the difference between their individual amounts of satisfaction. Unlike the EGALITARIAN STABLE MARRIAGE, the SEX-EQUAL STABLE MARRIAGE problem is known to be NP-hard [13]. On the other hand, in BALANCED STABLE MARRIAGE, the objective is to find a stable matching that minimizes $\mathsf{balance}(\mu) = \max\{\sum_{(m,w)\in\mu} p_m(w), \sum_{(m,w)\in\mu} p_w(m)\}$ over SM. At first sight, this measure might seem conceptually similar to the previous one, but in fact, the two measures are quite different. Indeed, BALANCED STABLE MARRIAGE does not attempt to find a stable matching that is fair, but one that is desirable by both sides. In other words, BALANCED STABLE MARRIAGE examines the amount of dissatisfaction of each party *individually*, and attempts to minimize the worse one among the two. This problem fits the common scenario in economics where each party is selfish in the sense that it desires a matching where its own dissatisfaction is minimized, irrespective of the dissatisfaction of the other party, and our goal is to find a matching desirable by both

parties by ensuring that each individual amount of dissatisfaction does not exceed some threshold. In some situations, the minimization of balance(μ) may indirectly also minimize $\delta(\mu)$, but in other situations, this may not be the case. Indeed, McDermid [17] constructed a family of instances where there does *not* exist any matching that is both a sex-equal stable matching and a balanced stable matching (the construction is also available in the book [16]).

The Balanced Stable Marriage problem was introduced in the influential work of Feder [6] on stable matchings. Feder [6] proved that this problem is NP-hard and that it admits a 2-approximation algorithm. Later, it was shown that this problem also admits a $(2 - 1/\ell)$-approximation algorithm where ℓ is the maximum size of a set of acceptable partners [16]. O'Malley [19] phrased the Balanced Stable Marriage problem in terms of constraint programming. Recently, McDermid and Irving [18] expressed interest in the design of fast exact exponential-time algorithms for Balanced Stable Marriage. In this paper, we study this problem in the realm of fast exact exponential-time algorithms as defined by the field of Parameterized Complexity (see Section 2). Recall that SM is the set of all stable matchings. In this context, we would like to remark that Egalitarian Stable Roommates problem is NP-complete [6]. Recently, Chen et al. [1] showed that it is FPT parameterized by the egalitarian cost. McDermid and Irving [18] showed that Sex-Equal Stable Marriage is NP-hard even if it is only necessary to decide whether the target $\Delta = \min_{\mu \in \text{SM}} |\delta(\mu)|$ is 0 or not [18]. In particular, this means that Sex-Equal Stable Marriage is not only W[1]-hard with respect to Δ, but it is even paraNP-hard with respect to this parameter.[1] In the case of Balanced Stable Marriage, however, fixed-parameter tractability with respect to the target $\text{Bal} = \min_{\mu \in \text{SM}} \text{balance}(\mu)$ trivially follows from the fact that this value is lower bounded by $\max\{|M|, |W|\}$.[2]

Our Contribution. We introduce two "above-guarantee parameterizations" of Balanced Stable Marriage. To this end, we consider the minimum value O_M of the total dissatisfaction of men that can be realized by a stable matching, and the minimum value O_W of the total dissatisfaction of women that can be realized by a stable matching. Formally, $O_M = \sum_{(m,w) \in \mu_M} p_m(w)$, and $O_W = \sum_{(m,w) \in \mu_W} p_w(m)$, where μ_M and μ_W are the man-optimal and woman-optimal stable matchings, respectively. An input integer k would indicate that our objective is to decide whether $\text{Bal} \leq k$. In our first parameterization, the parameter is $k - \min\{O_M, O_W\}$, and in the second one, it is $k - \max\{O_M, O_W\}$. In other words, we would like to answer the following questions (recall that $\text{Bal} = \min_{\mu \in \text{SM}} \text{balance}(\mu)$).

Above-Min Balanced Stable Marriage (Above-Min BSM)
Input: An instance $(M, W, \mathcal{L}_M, \mathcal{L}_W)$ of Balanced Stable Marriage, and a non-negative integer k.
Question: Is $\text{Bal} \leq k$?
Parameter: $t = k - \min\{O_M, O_W\}$

Above-Max Balanced Stable Marriage (Above-Max BSM)
Input: An instance $(M, W, \mathcal{L}_M, \mathcal{L}_W)$ of Balanced Stable Marriage, and a non-negative integer k.
Question: Is $\text{Bal} \leq k$?
Parameter: $t = k - \max\{O_M, O_W\}$

Before stating our results, let us explain the choice of these parameterizations. Here, note that the best satisfaction the party of men can hope for is O_M, and the best satisfaction the party of women can hope for is O_W. First, consider the parameter $t = k - \min\{O_M, O_W\}$.

---

[1]If a parameterized problem cannot be solved in polynomial time even when the value of the parameter is a fixed constant (that is, independent of the input), then the problem is said to be paraNP-hard.

[2]We assume that any stable matching μ is perfect. We justify this assumption later.

Whenever we have a solution such that the amounts of satisfaction of *both* parties are *close enough* to the best they can hope for, this parameter is small. Indeed, the closer the satisfaction of both parties to the best they can hope for (which is exactly the case where both parties would find the solution desirable), the smaller the parameter is, and the smaller the parameter is, the faster a parameterized algorithm is. In other words, if there exists a solution that is desirable by both parties, our parameter is small.

In the parameterization above, as we take the *min* of $\{O_M, O_W\}$, we need the satisfaction of *both* parties to be close to optimal in order to have a small parameter. As we are able to show that BALANCED STABLE MARRIAGE is FPT with respect to this parameter, it is very natural to next examine the case where we take the *max* of $\{O_M, O_W\}$. In this case, the closer the satisfaction of *at least one* party to the best it can hope for, the smaller the parameter is. In other words, now the demand from a solution—while not changing the definition of a solution in any way, that is, a solution is still a stable matching μ minimizing $\mathsf{balance}(\mu)$—in order to have a small parameter is *weaker*. In the jargon of Parameterized Complexity, it is said that the parameterization by $t = k - \max\{O_M, O_W\}$ is "above a higher guarantee" than the parameterization by $t = k - \min\{O_M, O_W\}$, since it is *always* the case that $\max\{O_M, O_W\} \geq \min\{O_M, O_W\}$. Unfortunately, as we show in this paper, the parameterization by $k - \max\{O_M, O_W\}$ results in a problem that is W[1]-hard. Hence, the complexities of the two parameterizations behave very differently. We remark that in Parameterized Complexity, it is *not at all* the rule that when one takes an "above a higher guarantee" parameterization, the problem would suddenly become W[1]-hard, as can be evidenced by the most classical above guarantee parameterizations in this field, which are of the VERTEX COVER problem. For that problem, three above guarantee parameterizations were considered in [3, 10, 15, 20], each above a higher guarantee than the previous one that was studied, and each led to a problem that is FPT. In that context, unlike our case, it is still not clear whether the bar can be raised higher. Overall, our results accurately draw the line between tractability and intractability with respect to the target value in the context of two very natural, useful parameterizations.

Finally, to be more precise, we note that our work proves three main theorems:

- First, we prove (in Section 3) that ABOVE-MIN BSM admits a kernel where *the number of people is linear in t*. For this purpose, we introduce notions that might be of independent interest in the context of a "functional" variant of ABOVE-MIN BSM. Our kernelization algorithm consists of several phases, each simplifying a different aspect of ABOVE-MIN BSM, and shedding light on structural properties of the YES-instances of this problem. Note that this result already implies that ABOVE-MIN BSM is FPT.

- Second, we prove (in Section 4) that ABOVE-MIN BSM admits a parameterized algorithm whose running time is *single exponential in the parameter t*. This algorithm first builds upon our kernel, and then incorporates the method of bounded search trees.

- Third, we prove (in Section 5) that ABOVE-MAX BSM is W[1]-hard. This reduction is quite technical, and its importance lies in the fact that it rules out (under plausible complexity-theoretic assumptions) the existence of a parameterized algorithm for ABOVE-MAX BSM. Thus, we show that although ABOVE-MAX BSM seems quite similar to ABOVE-MIN BSM, in the realm of Parameterized Complexity, these two problems are completely different.

2   Preliminaries

Let f be a function $f : A \to B$. For a subset $A' \subseteq A$, we let $f|_{A'}$ denote the restriction of f to A'. That is, $f|_{A'} : A' \to B$, and $f|_{A'}(a) = f(a)$ for all $a \in A'$.

Throughout the paper, whenever the instance $\mathcal{I}$ of BALANCED STABLE MARRIAGE under discussion is not clear from context or we would like to put emphasis on $\mathcal{I}$, we add "$(\mathcal{I})$" to the

appropriate notation. For example, we use the notation $t(\mathcal{I})$ rather than t. When we would like to refer to the balance of a stable matching μ in a specific instance $\mathcal{I}$, we would use the notation $\mathsf{balance}_{\mathcal{I}}(\mu)$. A matching is called a *perfect matching* if it matches every person (to some other person).

While designing our kernelization algorithm (see "Parameterized Complexity"), we might be able to determine whether the input instance is a YES-instance or a NO-instance. For the sake of clarity, in the first case, we simply return YES, and in second case, we simply return NO. To properly comply with the definition of a kernel, the return of YES and NO should be interpreted as the return of a trivial YES-instance and a trivial NO-instance, respectively. Here, a trivial YES-instance can be the one in which $M = W = \emptyset$ and $k = 0$, where the only stable matching is the one that is empty and whose balance is 0, and a trivial NO-instance can be the one where $M = \{m\}$, $W = \{w\}$, $\mathcal{A}(m) = \{w\}$, $\mathcal{A}(w) = \{m\}$ and $k = 0$.

Parameterized Complexity. A *parameterization* of a problem is the association of an integer k with each input instance, which results in a *parameterized problem*. For our purposes, we need to recall three central notions that define the parameterized complexity of a parameterized problem. The first one is the notion of a *kernel*. Here, a parameterized problem is said to admit a *kernel* of size $f(k)$ for some function f that depends *only* on k if there exists a polynomial-time algorithm, called a *kernelization algorithm*, that translates any input instance into an equivalent instance of the same problem whose size is bounded by $f(k)$ and such that the value of the parameter does not increase. In case the function f is polynomial in k, the problem is said to admit a *polynomial kernel*. Hence, kernelization is a mathematical concept that aims to analyze the power of preprocessing procedures in a formal, rigorous manner. The second notion that we use is the one of *fixed-parameter tractability (*FPT*)*. Here, a parameterized problem Π is said to be FPT if there is an algorithm that solves it in time $f(k) \cdot |I|^{\mathcal{O}(1)}$, where $|I|$ is the size of the input and f is a function that depends only on k. Such an algorithm is called a *parameterized algorithm*. In other words, the notion of FPT signifies that it is not necessary for the combinatorial explosion in the running time of an algorithm for Π to depend on the input size, but it can be confined to the parameter k. Finally, we recall that Parameterized Complexity also provides tools to refute the existence of polynomial kernels and parameterized algorithms for certain problems (under plausible complexity-theoretic assumptions), in which context the notion of W[1]-hard is a central one. It is widely believed that a problem that is W[1]-hard is unlikely to be FPT, and we refer the reader to the books [2,5] for more information on this notion in particular, and on Parameterized Complexity in general. The notation $\mathcal{O}^*$ is used to hide factors polynomial in the input size.

Reduction Rule. To design our kernelization algorithm, we rely on the notion of a *reduction rule*. A reduction rule is a polynomial-time procedure that replaces an instance $(\mathcal{I}, k)$ of a parameterized problem Π by a new instance $(\mathcal{I}', k')$ of Π. Roughly speaking, a reduction rule is useful when the instance $\mathcal{I}'$ is in some sense "simpler" than the instance $\mathcal{I}$. In particular, it is desirable to ensure that $k' \leq k$. The rule is said to be *safe* if $(\mathcal{I}, k)$ is a YES-instance if and only if $(\mathcal{I}', k')$ is a YES-instance.

A Functional Variant of STABLE MARRIAGE**.** To obtain our kernelization algorithm for ABOVE-MIN BSM, it will be convenient to work with a "functional" definition of preferences, resulting in a "functional" variant of this problem which we call ABOVE-MIN FBSM. Here, instead of the sets of preferences lists $\mathcal{L}_M$ and $\mathcal{L}_W$, the input consists of sets of preference *functions* $\mathcal{F}_M$ and $\mathcal{F}_W$, where $\mathcal{F}_M$ replaces $\mathcal{L}_M$ and $\mathcal{F}_W$ replaces $\mathcal{L}_W$. Specifically, every person $a \in M \cup W$ has an injective (one-to-one) function $f_a : \mathcal{A}(a) \to \mathbb{N}$, called a *preference function*. Intuitively, a lower function value corresponds to a higher preference. Since every preference function is injective, it defines a total order over a set of acceptable partners. Note that all of the definitions presented in the introduction extend to our functional variant in the natural way.

For the sake of formality, we specify the required adaptations below.

The input of the FUNCTIONAL STABLE MARRIAGE problem consists of a set of men, M, and a set of women, W. Each person a has a set of *acceptable partners*, denoted by $\mathcal{A}(a)$, and an injective function $f_a : \mathcal{A}(a) \to \mathbb{N}$ called a *preference function*. Without loss of generality, it is assumed that if a person a belongs to the set of acceptable partners of a person b, then the person b belongs to the set of acceptable partners of the person a. The set of preference functions of the men and the women are denoted by $\mathcal{F}_M$ and $\mathcal{F}_W$, respectively. A pair of a man and a woman, (m, a), is an *acceptable pair* if both $m \in \mathcal{A}(w)$ and $w \in \mathcal{A}(m)$. Accordingly, the notion of a *matching* refers to a matching between men and women, where two people that are matched to one another form an acceptable pair. A matching μ *stable* if it does not have a *blocking pair*, which is an acceptable pair (m, w) such that **(i)** either m is unmatched by μ or $f_m(w) < f_m(\mu(m))$, and **(ii)** either w is unmatched by μ or $f_w(m) < f_w(\mu(w))$. The goal of the FUNCTIONAL STABLE MARRIAGE problem is to find a stable matching.

The man-optimal stable matching, denoted by μ_M, is a stable matching such that every stable matching μ satisfies the following condition: every man m is either unmatched by both μ_M and μ or $f_m(\mu_M(m)) \leq f_m(\mu(m))$. The woman-optimal stable matching, denoted by μ_W, is defined analogously. Given a stable matching μ, define

$$\mathsf{balance}(\mu) = \max\{ \sum_{(m,w) \in \mu} f_m(w), \sum_{(m,w) \in \mu} f_w(m) \}.$$

Moreover, $\mathsf{Bal} = \min_{\mu \in \mathrm{SM}} \mathsf{balance}(\mu)$, where SM is the set of all stable matchings, $O_M = \sum_{(m,w) \in \mu_M} f_m(w)$, and $O_W = \sum_{(m,w) \in \mu_W} f_w(m)$. Finally, ABOVE-MIN FBSM is defined as follows.

ABOVE-MIN FUNCTIONAL BALANCED STABLE MARRIAGE (ABOVE-MIN FBSM)
Input: An instance $(M, W, \mathcal{F}_M, \mathcal{F}_W)$ of FUNCTIONAL BALANCED STABLE MARRIAGE, and a non-negative integer k.
Question: Is $\mathsf{Bal} \leq k$?
Parameter: $t = k - \min\{O_M, O_W\}$

From the above discussions it is straightforward to turn an instance of ABOVE-MIN BSM into an equivalent instance of ABOVE-MIN FBSM as stated in the following observation.

Observation 1. *Let $\mathcal{I} = (M, W, \mathcal{L}_M, \mathcal{L}_W, k)$ be an instance of* ABOVE-MIN BSM. *For each $a \in M \cup W$, define $f_a : \mathcal{A}(a) \to \mathbb{N}$ by setting $f_a(b) = p_a(b)$ for all $b \in \mathcal{A}(a)$. Then, $\mathcal{I}$ is a* YES-*instance of* ABOVE-MIN BSM *if and only if* $(M, W, \mathcal{F}_M = \{f_m\}_{m \in M}, \mathcal{F}_W = \{f_w\}_{w \in W}, k)$ *is a* YES-*instance of* ABOVE-MIN FBSM.

Known Results. Finally, we state several classical results, which were originally presented in the context of STABLE MARRIAGE. By their original proofs, these results also hold in the context of FUNCTIONAL STABLE MARRIAGE. To be more precise, given an instance of FUNCTIONAL STABLE MARRIAGE, we can construct an equivalent instance of STABLE MARRIAGE, by ranking the acceptable partners in the order of their function values, where a smaller value implies a higher preference. The instances are equivalent in the sense that they give rise to the exact same set of stable matchings. Hence, all the structural results about stable matchings in the usual setting (modeled by strict preference lists) apply to the generalized setting, modeled by injective functions.

Proposition 1 ([8]). *For any instance of* STABLE MARRIAGE *(or* FUNCTIONAL STABLE MARRIAGE*), there exist a man-optimal stable matching, μ_M, and a woman-optimal stable matching, μ_W, and both μ_M and μ_W can be computed in polynomial time.*

The following powerful proposition is known as the Rural-Hospital Theorem.

Proposition 2 ([9]). *Given an instance of* STABLE MARRIAGE *(or* ABOVE-MIN FBSM*), the set of men and women that are matched is the same for all stable matchings.*

We further need a proposition regarding the man-optimal and woman-optimal stable matchings that implies Proposition 2 [9].

Proposition 3 ([11]). *For any instance of* STABLE MARRIAGE *(or* FUNCTIONAL STABLE MARRIAGE*), every stable matching μ satisfies the following conditions: every woman w is either unmatched by both μ_M and μ or $p_w(\mu_M(w)) \geq p_w(\mu(w))$, and every man m is either unmatched by both μ_W and μ or $p_m(\mu_W(m)) \geq p_m(\mu(m))$.*

3 Kernel

In this section, we design a kernelization algorithm for ABOVE-MIN BSM. More precisely, we prove the following theorem.

Theorem 3.1. ABOVE-MIN BSM *admits a kernel that has at most $3t$ men, at most $3t$ women, and such that each person has at most $2t + 1$ acceptable partners.*

3.1 Functional Balanced Stable Marriage

To prove Theorem 3.1, we first prove the following result for the ABOVE-MIN FBSM problem.

Lemma 3.1. ABOVE-MIN FBSM *admits a kernel with at most $2t$ men, at most $2t$ women, and such that the image of the preference function of each person is a subset of $\{1, 2, \ldots, t+1\}$.*

To obtain the desired kernelization algorithm, we execute the following plan.

1. **Cleaning Prefixes and Suffixes.** Simplify the preference functions by "cleaning" suffixes and thereby also "cleaning" prefixes.
2. **Perfect Matching.** Zoom into the set of people matched by every stable matching.
3. **Overcoming Sadness.** Bound the number of "sad" people. Roughly speaking, a "sad" person a is one whose best attainable partner, b, does not reciprocate by considering a as the best attainable partner.
4. **Marrying Happy People.** Remove "happy" people from the instance.
5. **Truncating High-Values.** Obtain "compact" preference functions by truncating "high-values".
6. **Shrinking Gaps.** Shrink some of the gaps created by previous steps.

Each of the following subsections captures one of the steps above. In what follows, we let $\mathcal{I}$ denote our current instance of ABOVE-MIN FBSM. Initially, this instance is the input instance, but as the execution of our algorithm progresses, the instance is modified. The reduction rules that we present are applied *exhaustively* in the order of their presentation. In other words, at each point of time, the first rule whose condition is true is the one that we apply next. In particular, the execution terminates once the value of t drops below 0, as implied by the following rule.

Reduction Rule 1. *If $k < \max\{O_M, O_W\}$, then return* NO.

Lemma 3.3. *Reduction Rule 1 is safe.*

Proof. For every $\mu \in$ SM, it holds that $\mathsf{balance}(\mu) \geq \max\{O_M, O_W\}$. Thus, if $k < \max\{O_M, O_W\}$, then every $\mu \in$ SM satisfies $\mathsf{balance}(\mu) > k$. In this case, we conclude that $\mathsf{Bal} > k$, and therefore $\mathcal{I}$ is a NO-instance. □

Note that if $k < 0$, then it also holds that $t < 0$, and that if $t < 0$, then $k < \min\{O_M, O_W\}$. We remark that by Proposition 1, it would be clear that each of our reduction rules can indeed be implemented in polynomial time.

Cleaning Prefixes and Suffixes. We begin by modifying the images of the preference functions. We remark that it is *necessary* to perform this step first as otherwise the following steps would not be correct. To clean prefixes while ensuring *both* safeness and that the parameter t does not increase, we would actually need to clean *suffixes* first. Formally, we define suffixes as follows.

Definition 3.1. *Let (m, w) denote an acceptable pair. If m is matched by μ_W and $f_m(w) > f_m(\mu_W(m))$, then we say that w belongs to the suffix of m. Similarly, if w is matched by μ_M and $f_w(m) > f_w(\mu_M(w))$, then we say that m belongs to the suffix of w.*

By Proposition 3, we have the following observation.

Observation 2. *Let (m, w) denote an acceptable pair such that one of its members belongs to the suffix of the other member. Then, there is no $\mu \in \mathrm{SM}(\mathcal{I})$ that matches m with w.*

For every person a, let $\mathsf{worst}(a)$ be the person in $\mathcal{A}(a)$ to whom f_a assigns its worst preference value. More precisely, $\mathsf{worst}(a) = \mathrm{argmax}_{b \in \mathcal{A}(a)} f_a(b)$. We will now clean suffixes.

Reduction Rule 2. *If there exists a person a such that $\mathsf{worst}(a)$ belongs to the suffix of a, then define the preference functions as follows.*

- $f'_a = f_a|_{\mathcal{A}(a) \setminus \{\mathsf{worst}(a)\}}.$

- $f'_{\mathsf{worst}(a)} = f_{\mathsf{worst}(a)}|_{\mathcal{A}(\mathsf{worst}(a)) \setminus \{a\}}.$

- *For all $b \in M \cup W \setminus \{a, \mathsf{worst}(a)\}$: $f'_b = f_b$.*

The new instance is $\mathcal{J} = (M, W, \{f'_{m'}\}_{m' \in M}, \{f'_{w'}\}_{w' \in W}, k)$.

Lemma 3.4. *Reduction Rule 2 is safe, and $t(\mathcal{I}) = t(\mathcal{J})$.*

Proof. By the definition of the new preference functions, we have that for every $\mu \in \mathrm{SM}(\mathcal{I}) \cap \mathrm{SM}(\mathcal{J})$, it holds that $\sum_{(m,w) \in \mu} f_m(w) = \sum_{(m,w) \in \mu} f'_m(w)$ and $\sum_{(m,w) \in \mu} f_w(m) = \sum_{(m,w) \in \mu} f'_w(m)$. In particular, this means that to conclude that $\mathsf{Bal}(\mathcal{I}) = \mathsf{Bal}(\mathcal{J})$ (which implies safeness) as well as that $O_M(\mathcal{I}) = O_M(\mathcal{J})$ and $O_W(\mathcal{I}) = O_W(\mathcal{J})$ (which implies that $t(\mathcal{I}) = t(\mathcal{J})$), it is sufficient to show that $\mathrm{SM}(\mathcal{I}) = \mathrm{SM}(\mathcal{J})$. For this purpose, first consider some $\mu \in \mathrm{SM}(\mathcal{I})$. By Observation 2, it holds that $(a, \mathsf{worst}(a)) \notin \mu$. Hence, μ is a matching in $\mathcal{J}$. Moreover, if μ has a blocking pair in $\mathcal{J}$, then by the definition of the new preference functions, it is also a blocking pair in $\mathcal{I}$. Since μ is stable in $\mathcal{I}$, we have that $\mu \in \mathrm{SM}(\mathcal{J})$.

In the second direction, consider some $\mu \in \mathrm{SM}(\mathcal{J})$. Then, it is clear that μ is a matching in $\mathcal{I}$. Moreover, if μ has a blocking pair (m, w) in $\mathcal{I}$ that is not $(a, \mathsf{worst}(a))$, then (m, w) is an acceptable pair in $\mathcal{J}$, and therefore by the definition of the new preference functions, we have that (m, w) is also a blocking pair in $\mathcal{J}$. Hence, since μ is stable in $\mathcal{J}$, the only pair that can block μ in $\mathcal{I}$ is $(a, \mathsf{worst}(a))$. Thus, to show that $\mu \in \mathrm{SM}(\mathcal{I})$, it remains to prove that $(a, \mathsf{worst}(a))$ cannot block μ in $\mathcal{I}$. Suppose, by way of contradiction, that $(a, \mathsf{worst}(a))$ blocks μ in $\mathcal{I}$. In particular, this means that $f_a(\mathsf{worst}(a)) < f_a(\mu(a))$. However, this contradicts the definition of $\mathsf{worst}(a)$. $\square$

By cleaning suffixes, we actually also accomplish the objective of cleaning prefixes, which are defined as follows.

Definition 3.2. *Let (m, w) denote an acceptable pair. If m is matched by μ_M and $f_m(w) < f_m(\mu_M(m))$, then we say that w belongs to the prefix of m. Similarly, if w is matched by μ_W and $f_w(m) < f_w(\mu_W(w))$, then we say that m belongs to the prefix of w.*

Let us now claim that we have indeed succeeded in cleaning prefixes.

Lemma 3.5. *Let $\mathcal{I}$ be an instance of* ABOVE-MIN FBSM *on which Reduction Rules 1 to 2 have been exhaustively applied. Then, there does not exist an acceptable pair (m, w) such that one of it members belongs to the prefix of the other one.*

Proof. Suppose, by way of contradiction, that there exists an acceptable pair (m, w) such that one of its members belongs to the prefix of the other one. Without loss of generality, assume that w belongs to the prefix of m. Then, $f_m(w) < f_m(\mu_M(m))$. Since μ_M is a stable matching, it cannot be blocked by (m, w), which means that w is matched by μ_M and $f_w(\mu_M(w)) < f_w(m)$. Thus, we have that m belongs to the suffix of w, which contradicts the assumption that Reduction Rule 2 was applied exhaustively. $\square$

As a direct result of this lemma, we have the following corollary.

Corollary 3.1. *Let $\mathcal{I}$ be an instance of* ABOVE-MIN FBSM *on which Reduction Rules 1 to 2 have been exhaustively applied. Then, for every acceptable pair (m, w) in $\mathcal{I}$ where m and w are matched (not necessarily to each other) by both μ_M and μ_W, it holds that $f_m(\mu_M(m)) \leq f_m(w) \leq f_m(\mu_W(m))$ and $f_w(\mu_W(w)) \leq f_w(m) \leq f_w(\mu_M(w))$.*

Perfect Matching. Having Corollary 3.1 at hand, we are able to provide a simple rule that allows us to assume that every solution matches all people.

Reduction Rule 3. *If there exists a person unmatched by μ_M, then let M' and W' denote the subsets of men and women, respectively, who are matched by μ_M. For each $a \in M' \cup W'$, denote $\mathcal{A}'(a) = \mathcal{A}(a) \cap (M' \cup W')$, and define $f'_v = f_v|_{\mathcal{A}'(v)}$. The new instance is $\mathcal{J} = (M', W', \{f'_m\}_{m \in M'}, \{f'_w\}_{w \in W'}, k)$.*

To prove the safeness of this rule, we first prove the following lemma.

Lemma 3.6. *Let $\mathcal{I}$ be an instance of* ABOVE-MIN FBSM *on which Reduction Rules 1 to 2 have been exhaustively applied. Then, for every person a not matched by μ_M, it holds that $\mathcal{A}(a) = \emptyset$.*

Proof. Let a be a person not matched by μ_M. Then, by Proposition 2, it holds that a is not matched by any stable matching. Hence, we can assume w.l.o.g. that a is man m. First, note that $\mathcal{A}(m)$ cannot contain any woman w that is not matched by some stable matching, else (m, w) would have formed a blocking pair for that stable matching. Second, we claim that $\mathcal{A}(m)$ cannot contain a woman w that is matched by some stable matching. Suppose, by way of contradiction, that this claim is false. Then, by Proposition 2, it holds that $\mathcal{A}(m)$ contains a woman w that is matched by μ_M. We have that w prefers $\mu_M(w)$ over m, else (m, w) would have formed a blocking pair for μ_M, which is impossible as μ_M is a stable matching. However, this implies that m belongs to the suffix of w, which contradicts the supposition that Reduction Rule 2 has been exhaustively applied. We thus conclude that $\mathcal{A}(a) = \emptyset$. $\square$

Lemma 3.7. *Reduction Rule 3 is safe, and $t(\mathcal{I}) = t(\mathcal{J})$.*

Proof. By the definition of the new preference functions, we have that for every $\mu \in \mathrm{SM}(\mathcal{I}) \cap \mathrm{SM}(\mathcal{J})$, it holds that $\sum_{(m,w)\in\mu} f_m(w) = \sum_{(m,w)\in\mu} f'_m(w)$ and $\sum_{(m,w)\in\mu} f_w(m) = \sum_{(m,w)\in\mu} f'_w(m)$. To conclude that the lemma is correct, it is thus sufficient to argue that $\mathrm{SM}(\mathcal{I}) = \mathrm{SM}(\mathcal{J})$. For this purpose, first consider some $\mu \in \mathrm{SM}(\mathcal{I})$. By Proposition 2, we have that μ is also a matching in $\mathcal{J}$. Moreover, if μ has a blocking pair in $\mathcal{J}$, then by the definition of the new preference functions, it is also a blocking pair in $\mathcal{I}$. Since μ is stable in $\mathcal{I}$, we have that $\mu \in \mathrm{SM}(\mathcal{J})$.

In the second direction, consider some $\mu \in \mathrm{SM}(\mathcal{J})$. Then, it is clear that μ is a matching in $\mathcal{I}$. By Lemma 3.6, if μ has a blocking pair (m, w) in $\mathcal{I}$, then both $m \in M'$ and $w \in W'$. However, for such a blocking pair (m, w), we have that (m, w) is an acceptable pair in $\mathcal{J}$, and therefore by the definition of the new preference functions, we have that (m, w) is also a blocking pair in $\mathcal{J}$. Hence, since μ is stable in $\mathcal{J}$, we conclude that μ is also stable in $\mathcal{I}$. $\square$

By Proposition 2, from now onwards, we have that for the given instance, any stable matching is a perfect matching. Due to this observation, we can denote $n = |M| = |W|$, and for any stable matching μ, we have the following equalities.

$$\sum_{(m,w)\in\mu} f_m(w) = \sum_{m\in M} f_m(\mu(m)); \qquad \sum_{(m,w)\in\mu} f_w(m) = \sum_{w\in W} f_w(\mu(w)). \qquad \text{(I)}$$

Overcoming Sadness. As every stable matching is a perfect matching, every person is matched by every stable matching, including the man-optimal and woman-optimal stable matchings. Thus, it is well defined to classify the people who do not have the same partner in the man-optimal and woman-optimal stable matchings as "sad". That is,

Definition 3.3. *A person $a \in M \cup W$ is* sad *if $\mu_M(a) \neq \mu_W(a)$.*

We let M_S and W_S denote the sets of sad men and sad women, respectively. People who are not sad are termed *happy*. Accordingly, we let M_H and W_H denote the sets of happy men and happy women, respectively. Note that $M_S = \emptyset$ if and only if $W_S = \emptyset$. Moreover, note that by the definition of μ_M and μ_W, for a happy person a it holds that a and $\mu_M(a) = \mu_W(a)$ are matched to one another by every stable matching. Let us now bound the number of sad people in a YES-instance.

Reduction Rule 4. *If $|M_S| > 2t$ or $|W_S| > 2t$, then return* NO.

Lemma 3.8. *Reduction Rule 4 is safe.*

Proof. We only prove that if $|M_S| > 2t$, then $\mathcal{I}$ is a NO-instance, as the proof of the other case is symmetric to this one. Let us assume that $|M_S| > 2t$. Suppose, by way of contradiction, that $\mathcal{I}$ is a YES-instance. Then, there exists a stable matching μ such that $\mathsf{balance}(\mu) \leq k$. Partition $M_S = M'_S \uplus M''_S$ as follows. Set M'_S to be the set of all m in M_S such that m is not the partner of $\mu(m)$ in μ_W.

$$M'_S = \{m \in M_S \mid f_{\mu(m)}(m) > f_{\mu(m)}(\mu_W(\mu(m)))\}.$$

Accordingly, set $M''_S = M_S \setminus M'_S$. Since $|M_S| > 2t$, at least one among $|M'_S|$ and $|M''_S|$ is (strictly) larger than t. Let us first handle the case where $|M'_S| > t$. Then,

$$
\begin{aligned}
\sum_{w\in W} f_w(\mu(w)) &= \sum_{\{w\,|\,\mu(w)\in M'_S\}} f_w(\mu(w)) + \sum_{\{w\,|\,\mu(w)\notin M'_S\}} f_w(\mu(w)) \\
&\geq \sum_{\{w\,|\,\mu(w)\in M'_S\}} [f_w(\mu_W(w)) + 1] + \sum_{\{w\,|\,\mu(w)\notin M'_S\}} f_w(\mu_W(w)) \\
&= \sum_{w\in W} f_w(\mu_W(w)) + \sum_{\{w\,|\,\mu(w)\in M'_S\}} 1 = O_W + |M'_S| \\
&> O_W + t \geq k.
\end{aligned}
$$

Here, the first inequality followed directly from the definition of M'_S. As we have reached a contradiction, it must hold that $|M''_S| > t$. However, we now have that

$$
\begin{aligned}
\sum_{m\in M} f_m(\mu(m)) &= \sum_{m\in M''_S} f_m(\mu(m)) + \sum_{m\notin M''_S} f_m(\mu(m)) \\
&\geq \sum_{m\in M''_S} [f_m(\mu_M(m)) + 1] + \sum_{m\notin M''_S} f_m(\mu_M(m)) \\
&= \sum_{m\in M} f_m(\mu_M(m)) + \sum_{\{m\,|\,\mu(m)\in M''_S\}} 1 = O_M + |M''_S|
\end{aligned}
$$

$$> \quad O_M + t \geq k.$$

Here, the first inequality followed from the definition of M_S''. Indeed, for all $m \in M_S''$, we have that $f_{\mu(m)}(m) \leq f_{\mu(m)}(\mu_W(\mu(m)))$, else m would have belonged to M_S'. However, in this case we deduce that $m = \mu_W(\mu(m))$, and since $m \in M_S$, we have that $\mu(m) \neq \mu_M(m)$, which implies that $f_m(\mu(m)) \geq f_m(\mu_M(m)) + 1$. As we have again reached a contradiction, we conclude the proof. $\qquad\square$

Marrying Happy People. Towards the removal of happy people, we first need to handle the special case where there are no sad people. In this case, there is exactly one stable matching, which is the man-optimal stable matching (that is equal, in this case, to the woman-optimal stable matching). This immediately implies the safeness of the following rule.

Reduction Rule 5. *If $M_S = W_S = \emptyset$, then return* YES *if* balance$(\mu_M) \leq k$ *and* No *otherwise.*

Observation 3. *Reduction Rule 5 is safe.*

We now turn to discard happy people. When we perform this operation, we need to ensure that the balance of the instance is preserved. More precisely, we need to ensure that $\mathsf{Bal}(\mathcal{I}) = \mathsf{Bal}(\mathcal{J})$, where $\mathcal{J}$ denotes the new instance resulting from the removal of some happy people. Towards this, we let (m_h, w_h) denote a *happy pair*, which is simply a pair of a happy man and happy woman who are matched to each other in every stable matching. Then, we redefine the preference functions in a manner that allows us to transfer the "contributions" of m_h and w_h from $\mathsf{Bal}(\mathcal{I})$ to $\mathsf{Bal}(\mathcal{J})$ via some sad man and woman. We remark that these sad people exist because Reduction Rule 5 does not apply. The details are as follows.

Reduction Rule 6. *If there exists a happy pair (m_h, w_h), then proceed as follows. Select an arbitrary sad man m_s and an arbitrary sad woman w_s. Denote $M' = M \setminus \{m_h\}$ and $W' = W \setminus \{w_h\}$. For each person $a \in M' \cup W'$, the new preference function $f_a' : \mathcal{A}(a) \setminus \{m_h, w_h\} \to \mathbb{N}$ is defined as follows.*

- **The preference function of m_s.** *For each $w \in \mathcal{A}(m_s) \setminus \{w_h\}$: $f_{m_s}'(w) = f_{m_s}(w) + f_{m_h}(w_h)$.*

- **The preference function of w_s.** *For each $m \in \mathcal{A}(w_s) \setminus \{m_h\}$: $f_{w_s}'(m) = f_{w_s}(m) + f_{w_h}(m_h)$.*

- *For each $w \in W' \setminus \{w_s\}$: $f_w' = f_w|_{M'}$.*

- *For each $m \in M' \setminus \{m_s\}$: $f_m' = f_m|_{W'}$.*

The new instance is $\mathcal{J} = (M', W', \{f_m'\}_{m \in M'}, \{f_w'\}_{w \in W'}, k)$.

Let us first state a lemma concerning the proof of the forward direction of the safeness of Reduction Rule 6.

Lemma 3.9. *Let $\mu \in \mathrm{SM}(\mathcal{I})$. Then, $\mu' = \mu \setminus \{(m_h, w_h)\}$ is a stable matching in $\mathcal{J}$ such that* balance$_{\mathcal{J}}(\mu') = $ balance$_{\mathcal{I}}(\mu)$.

Proof. We first show that $\mu' \in \mathrm{SM}(\mathcal{J})$. By Reduction Rule 3, it holds that μ is a perfect matching in $\mathcal{I}$. Since (m_h, w_h) is a happy pair, it is clear that $(m_h, w_h) \in \mu$, and therefore μ' is a perfect matching in $\mathcal{J}$. Let $(m, w) \notin \mu'$ be some acceptable pair in $\mathcal{J}$. Since $\mu \in \mathrm{SM}(\mathcal{I})$ and it is a perfect matching, it holds that $f_m(w) > f_m(\mu(m))$ or $f_w(m) > f_w(\mu(w))$. Let us consider these two possibilities separately.

- Suppose that $f_m(w) > f_m(\mu(m))$. If $m \neq m_s$, then $f_m'(w) = f_m(w)$ and $f_m'(\mu'(m)) = f_m(\mu(m))$, and therefore $f_m'(w) > f_m'(\mu'(m))$. Else, $f_m'(w) = f_m(w) + f_{m_h}(w_h)$ and $f_m'(\mu'(m)) = f_m(\mu(m)) + f_{m_h}(w_h)$, and therefore again $f_m'(w) > f_m'(\mu'(m))$.

11

- Suppose that $f_w(m) > f_w(\mu(w))$. Analogously to the previous case, we get that $f'_w(m) > f'_w(\mu'(w))$.

Since the choice of (m, w) was arbitrary, we conclude that μ' does not have a blocking pair in $\mathcal{J}$, and therefore $\mu' \in \mathrm{SM}(\mathcal{J})$. To show that $\mathsf{balance}_{\mathcal{J}}(\mu') = \mathsf{balance}_{\mathcal{I}}(\mu)$, note that

$$
\begin{aligned}
\mathsf{balance}_{\mathcal{J}}(\mu') &= \max\{\sum_{m \in M \setminus \{m_h\}} f'_m(\mu'(m)), \sum_{w \in W \setminus \{w_h\}} f'_w(\mu'(w))\} \\
&= \max\{f'_{m_s}(\mu'(m_s)) + \sum_{m \in M \setminus \{m_h, m_s\}} f'_m(\mu'(m)), \\
&\qquad\quad f'_{w_s}(\mu'(w_s)) + \sum_{w \in W \setminus \{w_h, w_s\}} f'_w(\mu'(w))\} \\
&= \max\{f_{m_s}(\mu(m_s)) + f_{m_h}(w_h) + \sum_{m \in M \setminus \{m_h, m_s\}} f_m(\mu(m)), \\
&\qquad\quad f_{w_s}(\mu(w_s)) + f_{w_h}(m_h) + \sum_{w \in W \setminus \{w_h, w_s\}} f_w(\mu(w))\} \\
&= \max\{\sum_{m \in M} f_m(\mu(m)), \sum_{w \in W} f_w(\mu(w))\} = \mathsf{balance}_{\mathcal{I}}(\mu).
\end{aligned}
$$

This concludes the proof. $\qquad\square$

To prove a lemma addressing the reverse direction, we first state the following observation, whose correctness follows directly from Corollary 3.1.

Observation 4. *Let $\mathcal{I}$ be an instance of* Above-Min FBSM *on which Reduction Rules 1 to 2 have been exhaustively applied. Then, for every happy pair (m_h, w_h), it holds that $\mathcal{A}(m_h) = \{w_h\}$ and $\mathcal{A}(w_h) = \{m_h\}$.*

Lemma 3.10. *Let $\mu' \in \mathrm{SM}(\mathcal{J})$. Then, $\mu = \mu' \cup \{(m_h, w_h)\}$ is a stable matching in $\mathcal{I}$ such that $\mathsf{balance}_{\mathcal{I}}(\mu) = \mathsf{balance}_{\mathcal{J}}(\mu')$.*

Proof. We first show that $\mu \in \mathrm{SM}(\mathcal{I})$. By Reduction Rule 3, it holds that μ' is a perfect matching in $\mathcal{J}$. Since (m_h, w_h) is a happy pair and by Observation 4, we have that μ is a perfect matching in $\mathcal{I}$ such that neither m_h nor w_h participate in any pair that blocks μ (if such a pair exists). Let $(m, w) \notin \mu'$ be some acceptable pair in $\mathcal{I}$ such that $m \neq m_h$ and $w \neq w_h$. Since $\mu' \in \mathrm{SM}(\mathcal{J})$ and it is a perfect matching, it holds that $f'_m(w) > f'_m(\mu'(m))$ or $f'_w(m) > f'_w(\mu'(w))$. Let us consider these two possibilities separately.

- Suppose that $f'_m(w) > f'_m(\mu(m))$. If $m \neq m_s$, then $f_m(w) = f'_m(w)$ and $f_m(\mu(m)) = f'_m(\mu'(m))$, and therefore $f_m(w) > f_m(\mu(m))$. Else, $f_m(w) = f'_m(w) - f_{m_h}(w_h)$ and $f_m(\mu(m)) = f'_m(\mu'(m)) - f_{m_h}(w_h)$, and therefore again $f_m(w) > f_m(\mu(m))$.

- Suppose that $f'_w(m) > f'_w(\mu'(w))$. Analogously to the previous case, we get that $f_w(m) > f_w(\mu(w))$.

Since the choice of (m, w) was arbitrary, we conclude that μ does not have a blocking pair in $\mathcal{I}$, and therefore $\mu \in \mathrm{SM}(\mathcal{I})$. To show that $\mathsf{balance}_{\mathcal{I}}(\mu) = \mathsf{balance}_{\mathcal{J}}(\mu')$, we follow the exact same argument as the one present in the proof of Lemma 3.9. $\qquad\square$

We now turn to justify the use of Reduction Rule 6.

Lemma 3.11. *Reduction Rule 6 is safe, and $t(\mathcal{I}) = t(\mathcal{J})$.*

Proof. By Lemmata 3.9 and 3.10, it holds that $\mathsf{Bal}(\mathcal{I}) = \mathsf{Bal}(\mathcal{J})$, and that both $\mathsf{balance}_{\mathcal{I}}(\mu_M(\mathcal{I})) = \mathsf{balance}_{\mathcal{J}}(\mu_M(\mathcal{J}))$ and $\mathsf{balance}_{\mathcal{I}}(\mu_W(\mathcal{I})) = \mathsf{balance}_{\mathcal{J}}(\mu_W(\mathcal{J}))$. As the argument k remained untouched, we have that Reduction Rule 6 is safe as well as that $t(\mathcal{I}) = t(\mathcal{J})$. ☐

Before we proceed to examine preference functions more closely, let us take a step back and prove the following result.

Lemma 3.12. *Given an instance $\mathcal{I}$ of* ABOVE-MIN *FBSM, one can exhaustively apply Reduction Rules 1 to 6 in polynomial time to obtain an instance $\mathcal{J}$ such that $t(\mathcal{J}) \leq t(\mathcal{I})$. All people in $\mathcal{J}$ are sad and matched by every stable matching, and there exist at most $2t$ men and at most $2t$ women.*

Proof. Notice that each rule among Reduction Rules 1 to 6 can be applied in polynomial time, and it either terminates the execution of the algorithm or shrinks the size of the instance. Hence, it is clear that the instance $\mathcal{J}$ is obtained in polynomial time, and the claim that $t(\mathcal{J}) \leq t(\mathcal{I})$ follows from Lemmata 3.7 and 3.11. Due to Reduction Rules 3 and 6, we have that all people in $\mathcal{J}$ are sad and matched by every stable matching. Thus, due to Reduction Rule 4, we also have that there exist at most $2t$ men and at most $2t$ women. ☐

Truncating High-Values. Up until now, we have bounded the number of people. However, the images of the preference functions can contain integers that are not bounded by a function polynomial in the parameter. Thus, even though the number of people is upper bounded by $4t$, the total size of the instance can be huge. Accordingly, in what follows, we need to process the images of the preference functions. Recall that we have already modified preference functions so that they would not contain irrelevant information in their *prefixes* and *suffixes*. Our current goal is to truncate "high-values"' of preference functions. To understand the intuition behind the rule we present next, suppose that there exists a stable matching μ and a man m such that $f_m(\mu(m)) > t + f_m(\mu_M(m))$. That is, μ matches m to a woman whose position is larger than the position of the woman with whom m is matched by μ_M by at least t units. Then, $\mathsf{balance}(\mu) \geq \sum_{m' \in M} f_{m'}(\mu(m')) > O_M + t \geq k$. Hence, irrespective of whether or not the current instance is a YES-instance, we know that μ is not a yes-certificate. We thus observe that we should delete all those acceptable pairs whose presence in any stable matching prevents its balance from being upper bounded by k. Formally,

Reduction Rule 7. *If there exists an acceptable pair (m, w) such that $f_m(w) > (k - O_M) + f_m(\mu_M(m))$ or $f_w(m) > (k - O_W) + f_w(\mu_W(w))$, then define the preference functions as follows:*

- $f'_m = f_m|_{\mathcal{A}(m) \setminus \{w\}}$.

- $f'_w = f_w|_{\mathcal{A}(w) \setminus \{m\}}$.

- *For all $a \in M \cup W \setminus \{m, w\}$: $f'_a = f_a$.*

The new instance is $\mathcal{J} = (M, W, \{f'_{m'}\}_{m' \in M}, \{f'_{w'}\}_{w' \in W}, k)$.

Lemma 3.13. *Reduction Rule 7 is safe, and $t(\mathcal{I}) \geq t(\mathcal{J})$.*

Proof. Without loss of generality, suppose that $f_m(w) > (k - O_M(\mathcal{I})) + f_m(\mu_M(m))$. Due to Reduction Rule 1, we have that $k - O_M(\mathcal{I}) \geq 0$, and therefore $f_m(w) > f_m(\mu_M(m))$, which implies that $w \neq \mu_M(m)$. We thus have that $\mu_M(\mathcal{I})$ is also a matching in $\mathcal{J}$, and due to Corollary 3.1, we deduce that $\mu_M(\mathcal{I}) = \mu_M(\mathcal{J})$. First, we would like to show that $t(\mathcal{I}) \geq t(\mathcal{J})$. For this purpose, it is sufficient to show that $O_M(\mathcal{I}) \leq O_M(\mathcal{J})$ and $O_W(\mathcal{I}) \leq O_W(\mathcal{J})$. Since $\mu_M(\mathcal{I}) = \mu_M(\mathcal{J})$, it is clear that $O_M(\mathcal{I}) = O_M(\mathcal{J})$. By Reduction Rule 3, $\mu_M(\mathcal{I})$ matches all people in $\mathcal{I}$. Thus, by Proposition 2 and since $\mu_M(\mathcal{I}) = \mu_M(\mathcal{J})$, we have that $\mu_W(\mathcal{J})$ matches all people in $\mathcal{J}$, and hence all people in $\mathcal{I}$. By Corollary 3.1, for any woman w and the man m

most preferred by w in $\mathcal{J}$, it holds that w does not prefer m over $\mu_W(w)$ in $\mathcal{I}$. Thus, by the definition of the preference functions, we have that $O_W(\mathcal{I}) \leq O_W(\mathcal{J})$.

To show that the rule is safe, we need to show that $\mathsf{Bal}(\mathcal{I}) \leq k$ if and only if $\mathsf{Bal}(\mathcal{J}) \leq k$. For this purpose, let us first suppose that $\mathsf{Bal}(\mathcal{I}) \leq k$. Then, there exists $\mu \in \mathrm{SM}(\mathcal{I})$ such that $\mathsf{balance}_{\mathcal{I}}(\mu) \leq k$. Notice that if $\mu(m) = w$, then since $f_m(w) > (k - O_M) + f_m(\mu_M(m))$ and by the equations in "Perfect Matching", we have that

$$
\begin{aligned}
k \;&\geq\; \mathsf{balance}_{\mathcal{I}}(\mu) \\
&\geq\; \sum_{m' \in M} f_{m'}(\mu(m')) \\
&\geq\; (k - O_M(\mathcal{I})) + \sum_{m' \in M} f_{m'}(\mu_M(m')) > (k - O_M(\mathcal{I})) + O_M(\mathcal{I}),
\end{aligned}
$$

which is a contradiction. Hence, $(m, w) \notin \mu$. Therefore, μ is a matching in $\mathcal{J}$. By the definition of the new preference functions, if μ has a blocking pair in $\mathcal{J}$, then this pair also blocks μ in $\mathcal{I}$. Since μ is stable in $\mathcal{I}$, we have that μ is also stable in $\mathcal{J}$. Now, by our definition of the new preference functions, we have that $\mathsf{balance}_{\mathcal{I}}(\mu) = \mathsf{balance}_{\mathcal{J}}(\mu)$. Since $\mathsf{balance}_{\mathcal{I}}(\mu) \leq k$, we thus conclude that $\mathsf{Bal}(\mathcal{J}) \leq k$.

In the second direction, suppose that $\mathsf{Bal}(\mathcal{J}) \leq k$. Then, there exists $\mu \in \mathrm{SM}(\mathcal{J})$ such that $\mathsf{balance}_{\mathcal{J}}(\mu) \leq k$. Clearly, μ is also a matching in $\mathcal{I}$. Moreover, by the definition of the preference functions, every acceptable pair in $\mathcal{I}$ that is also present in $\mathcal{J}$ cannot block μ in $\mathcal{I}$, else it would have also blocked μ in $\mathcal{J}$. Thus, if μ has a blocking pair in $\mathcal{I}$, then this pair must be (m, w). We claim that (m, w) cannot block μ in $\mathcal{I}$, which would imply that $\mu \in \mathrm{SM}(\mathcal{I})$. Suppose, by way of contradiction, that this claim is not true. Recall that we have already proved that $\mu_M(\mathcal{I}) = \mu_M(\mathcal{J})$. Let us denote $\mu_M = \mu_M(\mathcal{I})$. We have that μ matches m, which implies that $f_m(\mu(m)) > f_m(w)$. Since $f_m(w) > (k - O_M(\mathcal{I})) + f_m(\mu_M(m))$, we deduce that $f_m(\mu(m)) > (k - O_M(\mathcal{I})) + f_m(\mu_M(m))$. Furthermore, since $f'_m(\mu(m)) = f_m(\mu(m))$, $f'_m(\mu_M(m)) = f_m(\mu_M(m))$ and $O_M(\mathcal{I}) = O_M(\mathcal{J})$, we get that $f'_m(\mu(m)) > (k - O_M(\mathcal{J})) + f'_m(\mu_M(m))$. However, we then have that

$$
\begin{aligned}
k \;&\geq\; \mathsf{balance}_{\mathcal{J}}(\mu) \\
&\geq\; \sum_{m' \in M} f'_{m'}(\mu(m')) \\
&\geq\; (k - O_M(\mathcal{J})) + \sum_{m' \in M} f'_{m'}(\mu_M(m')) > (k - O_M(\mathcal{J})) + O_M(\mathcal{J}),
\end{aligned}
$$

which is a contradiction. Therefore, $\mu \in \mathrm{SM}(\mathcal{I})$. The definition of the new preference functions imply that $\mathsf{balance}_{\mathcal{I}}(\mu) = \mathsf{balance}_{\mathcal{J}}(\mu)$. Since $\mathsf{balance}_{\mathcal{J}}(\mu) \leq k$, we conclude that $\mathsf{Bal}(\mathcal{I}) \leq k$. $\square$

Shrinking Gaps. Currently, there might still exist a man m or a woman w such that $f_m(\mu_M(m)) > 1$ or $f_w(\mu_W(w)) > 1$, respectively. In the following rule, we would like to decrease some values assigned by the preference functions of such men and women in a manner that preserves equivalence.

Reduction Rule 8. *If there exist $m \in M$ and $w \in W$ such that $f_m(\mu_M(m)) > 1$ and $f_w(\mu_W(w)) > 1$, then define the preference functions as follows.*

- **The preference function of m.** *For all $w' \in \mathcal{A}(m)$: $f'_m(w') = f_m(w') - 1$.*

- **The preference function of w.** *For all $m' \in \mathcal{A}(w)$: $f'_w(m') = f_w(m') - 1$.*

- *For all $a \in M \cup W \setminus \{m, w\}$: $f'_a = f_a$.*

The new instance is $\mathcal{J} = (M, W, \{f'_{m'}\}_{m' \in M}, \{f'_{w'}\}_{w' \in W}, k - 1)$.

Lemma 3.14. *Reduction Rule 8 is safe, and $t(\mathcal{I}) = t(\mathcal{J})$.*

Proof. Let us first observe that the set of acceptable pairs in $\mathcal{I}$ is equal to the set of acceptable pairs of $\mathcal{J}$. Furthermore, for every person a, including the cases where this person is either m or w, any two acceptable partners b and b' of a that satisfy $f_a(b) < f_a(b')$ also satisfy $f'_a(b) < f'_a(b')$, and vice versa. Indeed, this observation follows directly from our definition of the new preference functions. In other words, if a person prefers some person over another in $\mathcal{I}$, then this person also has the same preference order in $\mathcal{J}$, and vice versa. We thus deduce that $\mathrm{SM}(\mathcal{I}) = \mathrm{SM}(\mathcal{J})$.

We proceed by claiming that for all $\mu \in \mathrm{SM}(\mathcal{I})$, we have that $\mathsf{balance}_{\mathcal{I}}(\mu) = \mathsf{balance}_{\mathcal{J}}(\mu) + 1$. Indeed, by the definition of the new preference functions and the equations in "Perfect Matching", we have that

$$
\begin{aligned}
\mathsf{balance}_{\mathcal{I}}(\mu) \;&= \max\{ \sum_{m' \in M} f_{m'}(\mu(m')), \sum_{w' \in W} f_{w'}(\mu(w')) \} \\
&= \max\{ f_m(\mu(m)) + \sum_{m' \in M \setminus \{m\}} f_{m'}(\mu(m')), f_w(\mu(w)) + \sum_{w' \in W \setminus \{w\}} f_{w'}(\mu(w')) \} \\
&= \max\{ f'_m(\mu(m)) + 1 + \sum_{m' \in M \setminus \{m\}} f'_{m'}(\mu(m')), f'_w(\mu(w)) + 1 + \sum_{w' \in W \setminus \{w\}} f'_{w'}(\mu(w')) \} \\
&= \max\{ \sum_{m' \in M} f'_{m'}(\mu(m')), \sum_{w' \in W} f'_{w'}(\mu(w')) \} + 1 = \mathsf{balance}_{\mathcal{J}}(\mu) + 1.
\end{aligned}
$$

Furthermore, the arguments above also show that $O_M(\mathcal{I}) = O_M(\mathcal{J}) + 1$ and $O_W(\mathcal{I}) = O_W(\mathcal{J}) + 1$. Hence, we have that $\mathsf{Bal}(\mathcal{I}) = \mathsf{Bal}(\mathcal{J}) + 1$. Since k was decreased by 1, we conclude that the rule is safe and that $t(\mathcal{I}) = t(\mathcal{J})$. $\qquad\square$

The usefulness of Reduction Rule 8 lies in the observation that after the exhaustive application of this rule, at least one of the two parties does not have any member without a person assigned 1 by his/her preference function. More precisely,

Observation 5. *Let $\mathcal{I}$ be an instance of* ABOVE-MIN FBSM *that is reduced with respect to Reduction Rules 1 to 8. Then, either **(i)** for every $m \in M$, we have that $f_m(\mu_M(m)) = 1$, or **(ii)** for every $w \in W$, we have that $f_w(\mu_W(w)) = 1$. In particular, either **(i)** $O_M = |M|$ or **(ii)** $O_W = |W|$.*

This concludes the description of our reduction rules. We are now ready to prove Lemma 3.1.

Proof of Lemma 3.1. Given an instance $\mathcal{I}$ of ABOVE-MIN FBSM, our kernelization algorithm exhaustively applies Reduction Rules 1 to 8, after which it outputs the resulting instance, $\mathcal{J}$, as the kernel. Notice that each rule among Reduction Rules 1 to 8 can be applied in polynomial time, and it either terminates the execution of the algorithm or shrinks the size of the instance. Hence, it is clear that the instance $\mathcal{J}$ is obtained in polynomial time. The claims that $\mathcal{J}$ and $\mathcal{I}$ are equivalent and that $t(\mathcal{J}) \leq t(\mathcal{I})$ follow directly from the lemmata that prove the safeness of each rule as well as argue with respect to the parameter. By Lemma 3.12, we also have that the instance contains at most $2t$ (sad) men and at most $2t$ (sad) women.

It remains to show that the image of the preference function of each person is a subset of $\{1, 2, \ldots, t + 1\}$. By Reduction Rule 7, every acceptable pair (m, w) satisfies $f_m(w) \leq (k - O_M) + f_m(\mu_M(m))$ and $f_w(m) \leq (k - O_W) + f_w(\mu_W(w))$. Moreover, for every acceptable pair (m, w), it holds that $f_m(\mu_M(m)) \leq O_M - (|M| - 1)$ and $f_w(\mu_W(w)) \leq O_W - (|W| - 1)$. Thus, every acceptable pair (m, w) satisfies $f_m(w) \leq k - (|M| - 1)$ and $f_w(m) \leq k - (|W| - 1)$. By Reduction Rule 3 and Observation 5, we have that $t = k - \min\{O_M, O_W\} = k - |M| = k - |W|$. Hence, we further conclude that every acceptable pair (m, w) satisfies $f_m(w) \leq t + 1$ and $f_w(m) \leq t + 1$. $\qquad\square$

3.2 Balanced Stable Marriage

Having proved Lemma 3.1, we have a kernel for ABOVE-MIN FBSM. We would like to employ this kernelization algorithm to design one for ABOVE-MIN BSM. For this purpose, we need to remove gaps from preference functions. Once we do this, we can view preference functions as preference lists and obtain the desired kernel. In what follows, we describe our kernelization algorithm for ABOVE-MIN BSM.

Let $\mathcal{K} = (M', W', \mathcal{L}_{M'}, \mathcal{L}_{W'}, k')$ be the input instance, which is an instance of ABOVE-MIN BSM. Our algorithm begins by applying the reduction given by Observation 1 to translate $\mathcal{K}$ into an instance $\mathcal{I}' = (M', W', \mathcal{F}_{M'}, \mathcal{F}_{W'}, k')$ of ABOVE-MIN FBSM. Then, our algorithm applies the kernelization algorithm given by Lemma 3.1 to $\mathcal{I}'$, obtaining a reduced instance $\mathcal{I} = (M, W, \mathcal{F}_M, \mathcal{F}_W, k)$ of ABOVE-MIN FBSM. By Lemma 3.1, this instance has at most $2t$ men, at most $2t$ women, and the image of the preference function of each person is a subset of $\{1, 2, \ldots, t+1\}$. To eliminate "gaps" in the preference functions, the algorithm proceeds as described below. Note that we no longer apply any reduction rule from Section 3.1 (even if its condition is satisfied), as we currently give a *new* kernelization procedure rather than an extension of the previous one. Let us first formally define the notion of a gap.

Definition 3.4. *Let $a \in M \cup W$, and i be positive integer outside the image of f. If there exists an integer $j > i$ that belongs to the image of f, then f_a is said to have a* gap *at i.*

Inserting Dummies. We have ensured that the largest number in the image of any preference function is at most $t + 1$. As every person is sad, it has at least two acceptable partners, and hence it has at most $t - 1 \leq t$ gaps. To handle the gaps of *all* people, we create a set of t dummy men and t dummy women. Our objective is to introduce these dummy people as acceptable partners for people who have gaps in their preference functions, such that the function values of the dummy people would fill the gaps. In the context of the following rule, note that currently there are no happy people, and hence this rule would be applied (only once).

Reduction Rule 9. *If there do not exist happy people, then let $X = \{x_1, x_2, \ldots, x_t\}$ denote a set of new (dummy) men, and $Y = \{y_1, y_2, \ldots, y_t\}$ denote a set of new (dummy) women. For each $i \in \{1, 2, \ldots, t\}$, initialize $\mathcal{A}(x_i) = \{y_i\}$, $\mathcal{A}(y_i) = \{x_i\}$ and $f_{x_i}(y_i) = f_{y_i}(x_i) = 1$. The new instance is $\mathcal{J} = (M \cup X, W \cup Y, \{f_m\}_{m \in M \cup X}, \{f_w\}_{w \in W \cup Y}, k + t)$.*

We note that for all $i \in \{1, 2, \ldots, t\}$, it holds that (x_i, y_i) is a happy pair.

Lemma 3.15. *Reduction Rule 9 is safe, and $t(\mathcal{I}) = t(\mathcal{J})$.*

Proof. For all $i \in \{1, 2, \ldots, t\}$, it holds that (x_i, y_i) is a happy pair, and therefore it is present in every stable matching in $\mathcal{J}$. By our definition of the new preference functions, it is clear that if μ is a stable matching in $\mathcal{I}$, then $\mu' = \mu \cup \{(x_1, y_1), \ldots, (x_t, y_t)\}$ is a stable matching in $\mathcal{J}$. Moreover, if μ' is a stable matching in $\mathcal{J}$, then $\mu = \mu' \setminus \{(x_1, y_1), \ldots, (x_t, y_t)\}$ is a stable matching in $\mathcal{I}$. Hence, since for all $i \in \{1, 2, \ldots, t\}$, it holds that $f_{x_i}(y_i) = f_{y_i}(x_i) = 1$, our definition of the new preference functions directly implies that $\mathsf{Bal}(\mathcal{I}) + t = \mathsf{Bal}(\mathcal{J})$, $O_M(\mathcal{I}) + t = O_M(\mathcal{J})$ and $O_W(\mathcal{I}) + t = O_W(\mathcal{J})$. Hence, $t(\mathcal{I}) = t(\mathcal{J})$, which concludes the proof. □

Reduction Rule 10. [Male version] *If there exists $m \in M$ such that f_m has a gap at some j, then select some $y_i \in Y \setminus \mathcal{A}(m)$, and set $\mathcal{A}'(m) = \mathcal{A}(m) \cup \{y_i\}$ and $\mathcal{A}'(y_i) = \mathcal{A}(y_i) \cup \{m\}$. The preference functions are defined as follows.*

- **The preference function of m.** $f'_m(y_i) = j$, and for all $a \in \mathcal{A}(m)$, $f'_m(a) = f_m(a)$.

- **The preference function of y_i.** $f'_{y_i}(m) = \max_{m' \in \mathcal{A}(y_i)} (f_{y_i}(m') + 1)$, and for all $a \in \mathcal{A}(y_i)$, $f'_{y_i}(a) = f_{y_i}(a)$.

- *For all $a \in (M \cup W) \setminus \{m, y_i\}$: $f'_a = f_a$.*

The new instance is $\mathcal{J} = (M, W, \{f'_{m'}\}_{m' \in M}, \{f'_{w'}\}_{w' \in W}, k)$.

Lemma 3.16. *Reduction Rule 10 is safe, and $t(\mathcal{I}) = t(\mathcal{J})$.*

Proof. The only modifications that are performed are the insertion of m intro the set of acceptable partners of y_i as the least preferred person, and the insertion of y_i into the set of acceptable partners of m in a location that previosuly contained a gap. Let us first observe that since $f_{x_i}(y_i) = f_{y_i}(x_i) = 1$ and $f'_{x_i}(y_i) = f'_{y_i}(x_i) = 1$, it holds that (x_i, y_i) is a happy pair in both $\mathcal{I}$ and $\mathcal{J}$. Hence, it is clear that $\mathrm{SM}(\mathcal{I}) = \mathrm{SM}(\mathcal{J})$, $O_M(\mathcal{I}) = O_M(\mathcal{J})$, $O_W(\mathcal{I}) = O_W(\mathcal{J})$, and that the balance of any stable matching in $\mathcal{I}$ is equal to its balance in $\mathcal{J}$. We thus conclude that the rule is safe and that $t(\mathcal{I}) = t(\mathcal{J})$. □

Analogously, we have a female version of Reduction Rule 10 where we fill a gap in the preference function of some woman $w \in W$. We do not repeat our arguments again, and straightaway state the following result, which follows directly from the safeness of Reduction Rule 9 and the male and female versions of Reduction Rule 10.

Lemma 3.17. ABOVE-MIN FBSM *admits a kernel that has at most $3t$ men among whom at most $2t$ are sad, at most $3t$ women among whom at most $2t$ are sad, and such that each happy person has at most $2t + 1$ acceptable partners and each sad person has at most $t + 1$ acceptable partners.*

Finally, we translate the kernel for ABOVE-MIN FBSM to an instance of ABOVE-MIN BSM as follows. For all $a \in M \cup W$ and $b \in \mathcal{A}(a)$, we set $p_a(b) = f_a(b)$. The new instance is $\mathcal{J} = (M, W, \{p_m\}_{m \in M}, \{f_w\}_{w \in W}, k)$. Clearly, we thus obtain an equivalent instance, which leads us to the following somewhat stronger version of Theorem 3.1, relevant to Appendix 4.

Lemma 3.18. ABOVE-MIN BSM *admits a kernel that has at most $3t$ men among whom at $2t$ are sad, at most $3t$ women among whom at most $2t$ are sad, and such that each happy person has at most $2t + 1$ acceptable partners and each sad person has at most $t + 1$ acceptable partners. Moreover, every stable matching in the kernel is a perfect matching.*

This concludes the proof of Theorem 3.1.

4 Parameterized Algorithm

In this section, we design a parameterized algorithm for ABOVE-MIN BSM. More precisely, we prove the following theorem.

Theorem 4.1. ABOVE-MIN BSM *can be solved in time $\mathcal{O}^*(8^t)$.*

As our algorithm is based on the method of bounded search trees, we first give a brief description of this technique.

4.1 Bounded Search Tree

The running time of an algorithm that uses bounded search trees can be analyzed as follows (see, e.g., [2]). Suppose that the algorithm executes a branching rule which has ℓ branching options (each leading to a recursive call with the corresponding parameter value), such that, in the ith branch option, the current value of the parameter decreases by b_i. Then, $(b_1, b_2, \ldots, b_\ell)$ is called the *branching vector* of this rule. We say that α is the *root* of $(b_1, b_2, \ldots, b_\ell)$ if it is the (unique) positive real root of $x^{b^*} = x^{b^* - b_1} + x^{b^* - b_2} + \cdots + x^{b^* - b_\ell}$, where $b^* = \max\{b_1, b_2, \ldots, b_\ell\}$. If $r > 0$ is the initial value of the parameter, and the algorithm (a) returns a result when (or before) the parameter is negative, and (b) only executes branching rules whose roots are bounded by a constant $c > 0$, then its running time is bounded by $\mathcal{O}^*(c^r)$.

4.2 Description of the Algorithm

Given an instance $\widehat{\mathcal{I}} = (\widehat{M}, \widehat{W}, \widehat{\mathcal{L}}_M, \widehat{\mathcal{L}}_W, \widehat{k})$ of ABOVE-MIN BSM, we begin by using the procedure given by Lemma 3.18 to obtain (in polynomial time) a kernel $\mathcal{I} = (M, W, \mathcal{L}_M, \mathcal{L}_W, k)$ of ABOVE-MIN BSM such that $\mathcal{I}$ has at most $3t$ men among whom at most $2t$ are sad, at most $3t$ women among whom at most $2t$ are sad. Let us denote the happy pairs in $\mathcal{I}$ by $(x_1, y_1), (x_2, y_2), \ldots, (x_h, y_h)$ for the appropriate $h \leq t$, and the set of sad men by M_S. Note that $|M_S| \leq 2t$.

We proceed by executing a loop where each iteration corresponds to a different subset $M' \subseteq M_S$. For a specific iteration, our goal is to determine whether there exists a stable matching μ such that the following conditions are satisfied:

- $\mathsf{balance}(\mu) \leq k$.

- For all $m \in M'$: $\mu(m) \neq \mu_M(m)$.

- For all $m \in M_S \setminus M'$: $\mu(m) = \mu_M(m)$.

A stable matching satisfying the conditions above (in the context of the current iteration) is said to be *valid*. We denote $r = k - O_M$, and observe that $r \leq t$.

Let us now consider some specific iteration. To determine whether there exists a valid stable matching, our plan is to execute a branching procedure, called `Branch`, which outputs every set S of pairs of a man in M' and a woman, such that the following conditions are satisfied.

1. Every man m in M' participates in exactly one pair (m, w) of S, and for that unique pair, it holds that $w \in \mathcal{A}(m)$ and $p_m(w) > p_m(\mu_M(m))$.

2. $\displaystyle\sum_{m \in M'} (p_m(\mu(m)) - p_m(\mu_M(m))) \leq r$.

The description of this procedure is given in the following subsection. Here, let us argue that by having this procedure, we can conclude the proof of the correctness of the algorithm. In the current iteration, we examine each set S in the outputted family of sets. Then, we check whether the pairs in S, together with $(x_1, y_1), (x_2, y_2), \ldots, (x_h, y_h)$ and every pair in $\{(m, \mu_M(m)) : m \in M'\}$ form a stable matching whose balance is at most k. If the answer is positive, then we terminate the execution and accept, which is clearly correct.

At the end, if we did not accept in any iteration, we reject. To see why this decision is correct, suppose that there exists a stable matching μ whose balance is at most k. In this case, due to our exhaustive search, there exists some iteration in which μ is also valid. In that iteration, associated with some $M' \subseteq M - S$, observe that the set of pairs $\{(m, \mu(m)) : m \in M'\}$ is one of the outputted sets of pairs. Indeed, the satisfaction of Condition 1 follows from the fact that μ is a stable matching satisfying the last two conditions of validity, and Condition 2 follows from the fact that μ satisfies the first condition of validity.

Let us denote by T the running of the procedure `Branch`. Then, the total running time of our algorithm is bounded by $\mathcal{O}^*(2^{|M_S|} \cdot T) = \mathcal{O}^*(4^t \cdot T)$, and therefore to derive the running time in Theorem 4.1, we will need to ensure that $T = \mathcal{O}^*(2^t)$. For this purpose, it is sufficient to ensure that $T = \mathcal{O}^*(2^r)$.

4.3 The Branching Procedure

We now present the description of the procedure `Branch` in the context of some set $M' \subseteq M_S$. Let us denote $M' = \{m_1, m_2, \ldots, m_p\}$ for the appropriate p.

Each call to our procedure is of the form `Branch`(i, S) where $i \in \{0, 1, \ldots, p+1\}$ and S is a set of pairs of a man in $\{m_1, m_2, \ldots, m_i\}$ and a woman, such that the following conditions are satisfied.

1. Every man m in $\{m_1, m_2, \ldots, m_i\}$ participates in exactly one pair (m, w) of S, and for that unique pair, it holds that $w \in \mathcal{A}(m)$ and $p_m(w) > p_m(\mu_M(m))$. We define μ_S as the *function* whose domain is $\{m_1, m_2, \ldots, m_i\}$ and which assigns to each man m in its domain the unique woman w such that $(m, w) \in S$.

2. Define $\mathsf{balance}(S) = \displaystyle\sum_{m \in \{m_1, m_2, \ldots, m_i\}} (p_m(\mu_S(m)) - p_m(\mu_M(m)))$. Then, $\mathsf{balance}(S) \leq r$.

Note that in case $i = 0$, we have that $S = \emptyset$. The call to $\mathtt{Branch}$ that is performed by the algorithm given in Section 4.2 is precisely with these arguments, that is, $\mathtt{Branch}(0, \emptyset)$. The objective of a call $\mathtt{Branch}(i, S)$ is to return a family $\mathcal{F}$ of sets, where each set $F \in \mathcal{F}$ is a set of pairs of a man in $\{m_{i+1}, m_{i+2}, \ldots, m_p\}$ and a woman, such that the following conditions are satisfied.

1. Every man m in $\{m_{i+1}, m_{i+2}, \ldots, m_p\}$ participates in exactly one pair (m, w) of F, and for that unique pair, it holds that $w \in \mathcal{A}(m)$ and $p_m(w) > p_m(\mu_M(m))$. We define μ_F as the *function* whose domain is $\{m_{i+1}, m_{i+2}, \ldots, m_p\}$ and which assigns to each man m in its domain the unique woman w such that $(m, w) \in F$.

2. $\mathsf{balance}(S) + \displaystyle\sum_{m \in \{m_{i+1}, m_{i+2}, \ldots, m_p\}} (p_m(\mu_F(m)) - p_m(\mu_M(m))) \leq r$.

Clearly, by accomplishing this goal, we also achieve the objective posed in Section 4.2. The measure that we employ to analyze our procedure is $(r - \mathsf{balance}(S))$ (recall that $\mathsf{balance}(S)$ is defined in Condition 2 of the specification of a call), which is initially equal to r. Hence, according to the method of bounded search trees (see Section 4.1), to derive the running time $\mathcal{O}^*(2^r)$, it is sufficient to ensure that $\mathtt{Branch}$ **(a)** returns a result when (or before) the measure r is negative, and **(b)** only executes branching rules whose roots are bounded by 2. When the measure r is negative, we should simply return $\mathcal{F} = \emptyset$, as there does not exist a set F satisfying the conditions above. Otherwise, when $i = p + 1$, we can simply return $\mathcal{F} = \{\emptyset\}$.

Let us now consider a call $\mathtt{Branch}(i, S)$ where $r \geq 0$ and $i \leq p$. Denote $\widetilde{W} = \{w \in \mathcal{A}(m_{i+1}) : p_{m_{i+1}}(w) > p_{m_{i+1}}(\mu_M(m_{i+1}))\}$. We further refine $\widetilde{W}$ be letting W^* denote the set of those r women in $\widetilde{W}$ who are the most preferred by m_{i+1}. In case there are no such r women since $|\widetilde{W}| < r$, we simply denote $W^* = \widetilde{W}$. Let us also denote $W^* = \{w_1, w_2, \ldots, w_q\}$ for the appropriate $q \leq r$. Then, our procedure executes r branches. At the jth branch, $\mathtt{Branch}$ calls itself recursively with $(i+1, S \cup \{(m_{i+1}, w_j)\})$. Eventually, $\mathtt{Branch}$ returns $\bigcup_{j=1}^{q} \{\{(m_{i+1}, w_j)\} \cup F_j : F \in \mathcal{F}_j\}$ where for all $j \in \{1, 2 \ldots, q\}$, we set $\mathcal{F}_j$ to be the family of sets of pairs that was returned by the recursive call of the jth branch. The correctness follows from the observation that the process that we execute is an exhaustive search. More precisely, if there exists a set F satisfying Conditions 1 and 2, then it must include exactly one of the pairs in $\{(m_{i+1}, w_j) : j \in \{1, 2, \ldots, q\}\}$. Now, let us observe that at the jth branch, the measure changes from $r - \mathsf{balance}(S)$ to $r - \mathsf{balance}(S \cup \{(m_{i+1}, w_j)\})$. By our definition of w_j, we have that $p_{m_{i+1}}(w_j) - p_{m_{i+1}}(\mu_M(m_{i+1}))) = j$. Hence, at the worst case, the branching vector is $(1, 2, \ldots, r)$. Since the root of such a branching vector upper bounded by 2, our proof is complete.

5 Hardness

In this section, we prove the following theorem.

Theorem 5.1. Above-Max BSM *is* W[1]-hard.

For this purpose, we consider the Clique problem, which is defined as follows.

The CLIQUE problem is known to be W[1]-hard [4]. Thus, to prove Theorem 5.1, it is sufficient to prove the following result.

Lemma 5.1. *Given an instance* $\mathcal{I} = (G = (V, E), k)$ *of* CLIQUE*, an equivalent instance* $\widehat{\mathcal{I}} = (M, W, \mathcal{L}_M, \mathcal{L}_W, \widehat{k})$ *of* ABOVE-MAX BSM *such that* $t = 6(k + \dfrac{k(k-1)}{2})$ *can be constructed in time* $\mathcal{O}(f(k) \cdot |\mathcal{I}|^{\mathcal{O}(1)})$ *for some function* f.

The rest of this section focuses on the proof of Lemma 5.1. To this end, we let $\mathcal{I} = (G = (V, E), k)$ be some instance of CLIQUE. In Section 5.1, we construct (in "FPT time") an instance $\widehat{\mathcal{I}} = (M, W, \mathcal{L}_M, \mathcal{L}_W, \widehat{k})$ of ABOVE-MAX BSM. This section also contains an informal explanation of the intuition underlying the construction. Then, in Section 5.2, we verify that the parameter t associated with $\widehat{\mathcal{I}}$ is equal to $6(k + \dfrac{k(k-1)}{2})$. Finally, in Section 5.3, we prove that the input instance $\mathcal{I}$ of CLIQUE and our instance $\widehat{\mathcal{I}}$ of ABOVE-MAX BSM are equivalent. In what follows, we select arbitrary orders on V and E, according to which we denote $V = \{v_1, v_2, \ldots, v_{|V|}\}$ and $E = \{e_1, e_2, \ldots, e_{|E|}\}$.

5.1 Reduction

First, to construct the sets M and W, we define three pairwise-disjoint subsets of M, called M_V, M_E and $\widetilde{M}$, and three pairwise-disjoint subsets of W, called W_V, W_E and $\widetilde{W}$. Then, we set $M = M_V \cup M_E \cup \widetilde{M} \cup \{m^*\}$ and $W = W_V \cup W_E \cup \widetilde{W} \cup \{w^*\}$, where m^* and w^* denote a new man and a new woman (which do not belong to the six sets defined previously), respectively.

- $M_V = \{m_v^i : v \in V, i \in \{1, 2\}\}$; $W_V = \{w_v^i : v \in V, i \in \{1, 2\}\}$.

- $M_E = \{m_e^i : e \in E, i \in \{1, 2\}\}$; $W_E = \{w_e^i : e \in E, i \in \{1, 2\}\}$.

- Let $\delta = 2(|V| + |E| + |V||E| + |V||E|^2) - k(4 + 4k + 2|E| + (k-1)|V||E|)$.
  Then, $\widetilde{M} = \{\widetilde{m}^i : i \in \{1, 2, \ldots, \delta\}\}$ and $\widetilde{W} = \{\widetilde{w}^i : i \in \{1, 2, \ldots, \delta\}\}$.

Note that $|M| = |W|$. We remark that in what follows, we assume w.l.o.g. that $\delta \geq 0$ and $|V| > k + \dfrac{k(k-1)}{2}$, else the size of the input instance $\mathcal{I}$ of CLIQUE is bounded by a function of k and can therefore, by using brute-force, be solved in FPT time.

Before we proceed, let us discuss the intuition behind the definition of the subsets of men and women above, and the definitions of the preference lists that will follow. Roughly speaking, each pair of men, m_v^1 and m_v^2, represents a vertex, and we aim to ensure that either both men will be matched to their best partners (in the man-optimal stable matching) or both men will be matched to other partners (where there would be only *one* choice for these other partners that preserves stability). Accordingly, we will guarantee that the choice of matching these two men to their best partners translates to *not* choosing the vertex they represent into the clique, and the other choice translates to choosing this vertex into the clique.

Now, having just the set M_V, we can encode selection of vertices into the clique, but we cannot ensure that the vertices we select indeed form a clique. For this purpose, we also have the set M_E which, in a manner similar to M_V, encodes selection of edges into the clique. By designing the instance in a way that the situation of the men in the man-optimal stable matching is significantly worse than that of the women in the women-optimal stable matching, we are able

to ensure that at most $2(k + \dfrac{k(k-1)}{2})$ men in $M_V \cup M_E$ will not be assigned their best partners (here, we exploit the condition that $\mathsf{balance}(\mu) \leq \widehat{k}$ for a solution μ). We remark that here the man m^* plays a crucial role—by using dummy men and women (in the sets $\widetilde{M}$ and $\widetilde{W}$) that prefer each other over all other people, we ensure that the situation of m^* is always "extremely bad" (from his viewpoint), while the situation of his partner, w^*, is always "excellent" (from her viewpoint).

At this point, we first need to ensure that the edges that we select indeed connect the vertices that we select. For this purpose, we carefully design our reduction so that when a pair of men representing some edge e obtain partners worse than those they have in the man-optimal stable matching, it must be that the men representing the endpoints of e have also obtained partners worse than those they have in the man-optimal stable matching, else stability will not be preserved—the partners of the men represting the endpoints of e will form blocking pairs together with the men representing e.

Finally we observe that we still need to ensure that among our $2(k + \dfrac{k(k-1)}{2})$ distinguished men in $M_V \cup M_E$, which are associated with $k + \dfrac{k(k-1)}{2}$ selected elements (vertices and edges), there will be exactly $2k$ distinguished men from M_V and exactly $k(k-1)$ distinguished men from M_E, which would mean we have chosen k vertices and $\dfrac{k(k-1)}{2}$ edges. For this purpose, we construct an instance where for the women, it is only somewhat "beneficial" that the men in M_V will not be matched to their best partners, but it is extremely beneficial that the men in M_E will not be matched to their best partners. This objective is achieved by carefully placing dummy men (from $\widetilde{M}$) in the preference lists of women in W_E. By again exploiting the condition that $\mathsf{balance}(\mu) \leq \widehat{k}$ for a solution μ, we are able to ensure that there would be at least $k(k-1)$ distinguished men from M_E.

Next, we proceed with the formal presentation of our reduction by defining p_m for every man $m \in M$, and thus constructing $\mathcal{L}_M$.

- For all $m_v^1 \in M_V$: $\mathcal{A}(m_v^1) = \{w_v^1, \widetilde{w}^1, \widetilde{w}^2, w_v^2\}$.

    - $p_{m_v^1}(w_v^1) = 1$; $p_{m_v^1}(\widetilde{w}^1) = 2$; $p_{m_v^1}(\widetilde{w}^2) = 3$; $p_{m_v^1}(w_v^2) = 4$.

- For all $m_v^2 \in M_V$: $\mathcal{A}(m_v^2) = \{w_v^2, w^1, w^2, w_v^1\}$.

    - $p_{m_v^2}(w_v^2) = 1$; $p_{m_v^2}(\widetilde{w}^1) = 2$; $p_{m_v^2}(\widetilde{w}^2) = 3$; $p_{m_v^2}(w_v^1) = 4$.

- For all $m_{\{u,v\}}^1 \in M_E$ where $u < v$:[3] $\mathcal{A}(m_{\{u,v\}}^1) = \{w_{\{u,v\}}^1, w_u^1, w_v^1, w_{\{u,v\}}^2\}$.

    - $p_{m_{\{u,v\}}^1}(w_{\{u,v\}}^1) = 1$; $p_{m_{\{u,v\}}^1}(w_u^1) = 2$; $p_{m_{\{u,v\}}^1}(w_v^1) = 3$; $p_{m_{\{u,v\}}^1}(w_{\{u,v\}}^2) = 4$.

- For all $m_{\{u,v\}}^2 \in M_E$ where $u < v$: $\mathcal{A}(m_{\{u,v\}}^2) = \{w_{\{u,v\}}^2, w_u^2, w_v^2, w_{\{u,v\}}^1\}$.

    - $p_{m_{\{u,v\}}^2}(w_{\{u,v\}}^2) = 1$; $p_{m_{\{u,v\}}^2}(w_u^2) = 2$; $p_{m_{\{u,v\}}^2}(w_v^2) = 3$; $p_{m_{\{u,v\}}^2}(w_{\{u,v\}}^1) = 4$.

- For all $\widetilde{m}^i \in \widetilde{M}$ such that $i \leq |V||E|$: $\mathcal{A}(\widetilde{m}^i) = \{\widetilde{w}^i\} \cup W_E \cup \{w_v^i \in M_V : \widetilde{m}^i \in \mathcal{A}(w_v^i)\}$, where for all $w_v^i \in W_V$, the set $\mathcal{A}(w_v^i)$ is determined later.

    - $p_{\widetilde{m}^i}(\widetilde{w}^i) = 1$.

---

[3]Recall that we have defined an order on V.

- For all $j \in \{1, 2, \ldots, |E|\}$: $p_{\widetilde{m}^i}(w_{e_j}^1) = j + 1$; $p_{\widetilde{m}^i}(w_{e_j}^2) = |E| + j + 1$.
- Denote $X = \{w_v^i \in M_V : \widetilde{m}^i \in \mathcal{A}(w_v^i)\}$. Let $f : X \to |X|$ be an arbitrarily chosen bijection. Then, for all $w_v^i \in X$: $p_{\widetilde{m}^i}(w_v^i) = 2|E| + f(w_v^i) + 1$.

- For all $\widetilde{m}^i \in \widetilde{M}$ such that $i > |V||E|$: $\mathcal{A}(\widetilde{m}^i) = \{\widetilde{w}^i\}$.

  - $p_{\widetilde{m}^i}(\widetilde{w}^i) = 1$.

- For m^*: $\mathcal{A}(m^*) = \widetilde{W} \cup \{w^*\}$.

  - For all $i \in \{1, 2, \ldots, \delta\}$: $p_{m^*}(\widetilde{w}^i) = i$.
  - $p_{m^*}(w^*) = \delta + 1$.

Accordingly, we define p_w for every woman $w \in W$, and thus construct $\mathcal{L}_W$.

- For all $w_v^1 \in W_V$: $\mathcal{A}(w_v^1) = \{m_e^1 \in M_E : v \in e\} \cup \{m_v^2, m_v^1\} \cup \{\widetilde{m}^i \in \widetilde{M} : i \leq |E| - \text{degree}_G(v)\}$.

  - $p_{w_v^1}(m_v^2) = 1$
  - For all $i \in \{1, 2, \ldots, |E|\}$: $p_{w_v^1}(m_{e_i}^1) = i + 1$.
  - $p_{w_v^1}(m_v^1) = |E| + 2$.

- For all $w_v^2 \in W_V$: $\mathcal{A}(w_v^2) = \{m_e^2 \in M_E : v \in e\} \cup \{m_v^1, m_v^2\} \cup \{\widetilde{m}^i \in \widetilde{M} : i \leq |E| - \text{degree}_G(v)\}$.

  - $p_{w_v^2}(m_v^1) = 1$.
  - For all $i \in \{1, 2, \ldots, |E|\}$: $p_{w_v^2}(m_{e_i}^2) = i + 1$.
  - $p_{w_v^2}(m_v^2) = |E| + 2$.

- For all $w_e^1 \in W_E$: $\mathcal{A}(w_e^1) = \{\widetilde{m}^i \in \widetilde{M} : i \in \{1, 2, \ldots, |V||E|\}\} \cup \{m_e^2, m_e^1\}$.

  - $p_{w_e^1}(m_e^2) = 1$.
  - For all $i \in \{1, 2, \ldots, |V||E|\}$: $p_{w_e^1}(\widetilde{m}^i) = i + 1$.
  - $p_{w_e^1}(m_e^1) = |V||E| + 2$.

- For all $w_e^2 \in W_E$: $\mathcal{A}(w_e^2) = \{\widetilde{m}^i \in \widetilde{M} : i \in \{1, 2, \ldots, |V||E|\}\} \cup \{m_e^1, m_e^2\}$.

  - $p_{w_e^2}(m_e^1) = 1$.
  - For all $i \in \{1, 2, \ldots, |V||E|\}$: $p_{w_e^2}(\widetilde{m}^i) = i + 1$.
  - $p_{w_e^2}(m_e^2) = |V||E| + 2$.

- For all $\widetilde{w}^i \in \widetilde{W}$ such that $i \in \{1, 2\}$: $\mathcal{A}(\widetilde{w}^i) = \{\widetilde{m}^i\} \cup \{m^*\} \cup M_V$.

  - $p_{\widetilde{w}^i}(\widetilde{m}^i) = 1$; $p_{\widetilde{w}^i}(m^*) = 2$.
  - For all $j \in \{1, 2, \ldots, |V|\}$: $p_{\widetilde{w}^i}(m_{v_j}^1) = j + 2$; $p_{\widetilde{w}^i}(m_{v_j}^2) = |V| + j + 2$.

- For all $\widetilde{w}^i \in \widetilde{W}$ such that $i > 2$: $\mathcal{A}(\widetilde{w}^i) = \{\widetilde{m}^i\} \cup \{m^*\}$.

  - $p_{\widetilde{w}^i}(\widetilde{m}^i) = 1$; $p_{\widetilde{w}^i}(m^*) = 2$.

- For w^*: $\mathcal{A}(w^*) = \{m^*\}$.
    - $p_{w^*}(m^*) = 1$.

Finally, we define $\widehat{k} = |M| + \delta + 6(k + \dfrac{k(k-1)}{2})$. It is clear that the entire construction (under the assumptions that $\delta \geq 0$ and $|V| > k + \dfrac{k(k-1)}{2}$) can be performed in polynomial time.

5.2 The Parameter

Our current objective is to verify that t is indeed bounded by a function of k. For this purpose, we first observe that for all $i \in \{1, 2, \ldots, \delta\}$, it holds that $p_{\widetilde{m}^i}(\widetilde{w}^i) = p_{\widetilde{w}^i}(\widetilde{m}^i) = 1$. Therefore, for all $\mu \in \mathrm{SM}(\widehat{\mathcal{I}})$ and $i \in \{1, 2, \ldots, \delta\}$, we have that $\mu(\widetilde{m}^i) = \widetilde{w}^i$, else $(\widetilde{m}^i, \widetilde{w}^i)$ would have formed a blocking pair for μ in $\widehat{\mathcal{I}}$.

Observation 6. *For all $\mu \in SM(\widehat{\mathcal{I}})$ and $i \in \{1, 2, \ldots, \delta\}$, it holds that $\mu(\widetilde{m}^i) = \widetilde{w}^i$.*

Now, note that $\mathcal{A}(m^*) = \widetilde{W} \cup \{w^*\}$. Thus, by Observation 6, we have that for all $\mu \in SM(\widehat{\mathcal{I}})$, either m^* is unmatched or $\mu(m^*) = w^*$. However, $\mathcal{A}(w^*) = \{m^*\}$, which implies that in the former case, (m^*, w^*) forms a blocking pair. Thus, we also have the following observation.

Observation 7. *For all $\mu \in SM(\widehat{\mathcal{I}})$, it holds that $\mu(m^*) = w^*$.*

Let us proceed by identifying the man-optimal μ_M and the woman-optimal μ_W stable matchings. For this purpose, we first define a matching μ'_M as follows.

- For all $m^i_v \in M_V$: $\mu'_M(m^i_v) = w^i_v$.
- For all $m^i_e \in M_E$: $\mu'_M(m^i_e) = w^i_e$.
- For all $\widetilde{m}^i \in \widetilde{M}$: $\mu'_M(\widetilde{m}^i) = \widetilde{w}^i$.
- $\mu'_M(m^*) = w^*$.

Lemma 5.2. $\mu_M = \mu'_M$.

Proof. Since for all $i \in \{1, 2, \ldots, \delta\}$, it holds that $\mu'_M(\widetilde{m}^i) = \widetilde{w}^i$ and $p_{\widetilde{m}^i}(\widetilde{w}^i) = p_{\widetilde{w}^i}(\widetilde{m}^i) = 1$, we have that there cannot exist a blocking pair with at least one person from $\widetilde{M} \cup \widetilde{W}$. Now, notice that for every $m \in M$, including m^*, the woman most preferred by m who is outside $\widetilde{W}$ is also the one with whom it is matched. Therefore, there cannot exist any blocking pair for μ'_M, and by Observation 6, we further conclude that indeed $\mu_M = \mu'_M$. $\square$

Now, we define a matching μ'_W as follows.

- For all $w^i_v \in W_V$: $\mu'_W(w^i_v) = m^{3-i}_v$.
- For all $w^i_e \in W_E$: $\mu'_W(w^i_e) = m^{3-i}_e$.
- For all $\widetilde{w}^i \in \widetilde{W}$: $\mu'_W(\widetilde{w}^i) = \widetilde{m}^i$.
- $\mu'_W(w^*) = m^*$.

Lemma 5.3. $\mu_W = \mu'_W$.

Proof. In the matching μ'_W, every woman is matched with the man she prefers the most. Thus, it is immediate that $\mu_W = \mu'_W$. $\square$

As a corollary to Lemmata 5.2 and 5.3, we obtain the following result.

Corollary 5.1. $O_M = |M| + \delta$ and $O_W = |W|$.

Proof. First, note that

$$O_M = \sum_{(m,w)\in\mu_M} p_m(w)$$
$$= p_{m^*}(\mu_M(m^*)) + \sum_{m\in M\setminus\{m^*\}} p_m(\mu_M(m)) = (\delta + 1) + (|M| - 1) = |M| + \delta.$$

Second, note that

$$O_W = \sum_{(m,w)\in\mu_W} p_w(m) = \sum_{w\in W} p_w(\mu_W(w)) = |W|.$$

$\square$

We are now ready to bound t.

Lemma 5.4. *The parameter t associated with $\widehat{\mathcal{I}}$ is equal to $6(k + \dfrac{k(k-1)}{2})$.*

Proof. By the definition of t, we have that

$$t = \widehat{k} - \max\{O_M, O_W\}$$
$$= |M| + \delta + 6(k + \frac{k(k-1)}{2}) - \max\{|M| + \delta, |W|\} = 6(k + \frac{k(k-1)}{2}).$$

$\square$

5.3 Correctness

First, from Lemma 5.2 and Proposition 2 we derive the following useful observation.

Observation 8. *Every $\mu \in SM(\widehat{\mathcal{I}})$ matches all people in $M \cup W$.*

Next, we proceed to state the first direction necessary to conclude that the input instance $\mathcal{I}$ of CLIQUE and our instance $\widehat{\mathcal{I}}$ of ABOVE-MAX BSM are equivalent.

Lemma 5.5. *If $\mathcal{I}$ is a YES-instance, then $\widehat{\mathcal{I}}$ is a YES-instance.*

Proof. Suppose that $\mathcal{I}$ is a YES-instance, and let U be the vertex-set of a clique on k vertices in G. We denote $M_V^U = \{m_v^i \in M_V : v \in U\}$ and $M_E^U = \{m_{\{u,v\}}^i \in M_E : u, v \in U\}$. Then, we define a matching μ as follows.

- For all $m_v^i \in M_V$:
  - If $m_v^i \in M_V^U$: $\mu(m_v^i) = w_v^{3-i}$.
  - Else: $\mu(m_v^i) = w_v^i$.

- For all $m_e^i \in M_E$:
  - If $m_e^i \in M_E^U$: $\mu(m_e^i) = w_e^{3-i}$.
  - Else: $\mu(m_e^i) = w_e^i$.

- For all $\widetilde{m}^i \in \widetilde{M}$: $\mu(\widetilde{m}^i) = \widetilde{w}^i$.

- $\mu(m^*) = w^*$.

We claim that $\mu \in SM(\widehat{\mathcal{I}})$ and $\mathsf{balance}(\mu) \leq \widehat{k}$, which would imply that $\widehat{\mathcal{I}}$ is a Yes-instance. To this end, we first show that $\mu \in SM(\widehat{\mathcal{I}})$. Since for all $i \in \{1, 2, \ldots, \delta\}$, it holds that $\mu(\widetilde{m}^i) = \widetilde{w}^i$ and $p_{\widetilde{m}^i}(\widetilde{w}^i) = p_{\widetilde{w}^i}(\widetilde{m}^i) = 1$, we have that there cannot exist a blocking pair with at least one person from $\widetilde{M} \cup \widetilde{W}$. Thus, there can also not be a blocking pair with any person from $\{m^*, w^*\}$.

On the one hand, notice that for every $m \in (M_V \setminus M_V^U) \cup (M_E \setminus M_E^U) \cup \{m^*\}$, the woman most preferred by m who is outside $\widetilde{W}$ is also the one with whom it is matched. Thus, no man in $(M_V \setminus M_V^U) \cup (M_E \setminus M_E^U) \cup \{m^*\}$ can belong to a blocking pair. Moreover, the set of acceptable partners of any woman in W_E matched to a man in $M_E \setminus M_E^U$ is a subset of $\widetilde{M} \cup (M_E \setminus M_E^U)$, and therefore such a woman cannot belong to a blocking pair. On the other hand, let W' denote the set of every woman that is matched to a man $m \in M_V^U \cup M_E^U$. Then, for every $w \in W'$, the man most preferred by w is also the one with whom she is matched. Therefore, no woman in W' can belong to a blocking pair. Hence, we also conclude that no woman in W_E can belong to a blocking pair.

Thus, if there exists a blocking pair, it must consist of a man $m \in M_V^U \cup M_E^U$ and a woman $w \in W_V \setminus W'$. Suppose, by way of contradiction, that there exists such a blocking pair (m, w). First, let us assume that $m = m_v^i \in M_V^U$. In this case, since apart from w_v^i, all women in $\mathcal{A}(m_v^i)$ belong to $\widetilde{W} \cup \{\mu(m_v^i)\}$, we deduce that $w = w_v^i$. However, w_v^i prefers $\mu(w_v^i)$ over m_v^i, and thus we reach a contradiction. Next, we assume that $m = m_{\{u,v\}}^i \in M_E^U$. In this case, it must hold that w is either w_v^i or w_u^i. Without loss of generality, we assume that $w = w_v^i$. However, since $m_{\{u,v\}}^i \in M_E^U$, we have that $v \in U$. Therefore, $\mu(w_v^i) = m_v^{3-i}$. Since w_v^i prefers m_v^{3-i} over $m_{\{u,v\}}^i$, we reach a contradiction.

It remains to prove that $\mathsf{balance}(\mu) \leq \widehat{k}$. To this end, we need to show that

$$\max\Big\{ \sum_{(m,w)\in\mu} p_m(w), \sum_{(m,w)\in\mu} p_w(m) \Big\} \leq |M| + \delta + 6(k + \frac{k(k-1)}{2}).$$

First, note that

$$
\begin{aligned}
\sum_{(m,w)\in\mu} p_m(w) &= \sum_{m\in M_V^U} p_m(\mu(m)) + \sum_{m\in M_V\setminus M_V^U} p_m(\mu(m)) + \sum_{m\in M_E^U} p_m(\mu(m)) \\
&\quad + \sum_{m\in M_E\setminus M_E^U} p_m(\mu(m)) + \sum_{m\in\widetilde{M}} p_m(\mu(m)) + p_{m^*}(\mu(m^*)) \\
&= 4|M_V^U| + |M_V\setminus M_V^U| + 4|M_E^U| + |M_E\setminus M_E^U| + |\widetilde{M}| + \delta + 1 \\
&= |M| + \delta + 3(|M_V^U| + |M_E^U|) = |M| + \delta + 6(k + \frac{k(k-1)}{2}).
\end{aligned}
$$

Second, note that

$$
\begin{aligned}
\sum_{(m,w)\in\mu} p_w(m) &= \sum_{m\in M_V^U} p_{\mu(m)}(m) + \sum_{m\in M_V\setminus M_V^U} p_{\mu(m)}(m) + \sum_{m\in M_E^U} p_{\mu(m)}(m) \\
&\quad + \sum_{m\in M_E\setminus M_E^U} p_{\mu(m)}(m) + \sum_{m\in\widetilde{M}} p_{\mu(m)}(m) + p_{\mu(m^*)}(m^*) \\
&= |M_V^U| + |M_V\setminus M_V^U|(|E|+2) + |M_E^U| + |M_E\setminus M_E^U|(|V||E|+2) + |\widetilde{M}| + 1 \\
&= |M| + 2(|V|-k)(|E|+1) + 2(|E| - \frac{k(k-1)}{2})(|V||E|+1) \\
&= |M| + \delta + 6(k + \frac{k(k-1)}{2}).
\end{aligned}
$$

This concludes the proof of the lemma. $\qquad\square$

We now turn to prove the second direction.

Lemma 5.6. *If $\widehat{\mathcal{I}}$ is a* Yes-*instance, then $\mathcal{I}$ is a* Yes-*instance.*

Proof. Suppose that $\widehat{\mathcal{I}}$ is a Yes-instance, and let μ be a stable matching such that $\mathsf{balance}(\mu) \leq \widehat{k}$. By Observations 6 and 7, it holds that

- For all $i \in \{1, 2, \ldots, \delta\}$: $\mu(\widetilde{m}^i) = \widetilde{w}^i$.

- $\mu(m^*) = w^*$.

Thus, since Observation 8 implies that all vertices in M_V should be matched by μ, we deduce that

- For all $v \in V$: Either both $\mu(m_v^1) = w_v^1$ and $\mu(m_v^2) = w_v^2$ or both $\mu(m_v^2) = w_v^1$ and $\mu(m_v^1) = w_v^2$.

Let U denote the set of every $v \in V$ such that $\mu(m_v^2) = w_v^1$ and $\mu(m_v^1) = w_v^2$. Moreover, denote $M_V^U = \{m_v^i \in M_V : v \in U\}$. By the item above, and since all vertices in M_E should also be matched by μ, we further deduce that

- For all $e \in E$: Either both $\mu(m_e^1) = w_e^1$ and $\mu(m_e^2) = w_e^2$ or both $\mu(m_e^2) = w_e^1$ and $\mu(m_e^1) = w_e^2$.

Let S denote the set of every $e \in E$ such that $\mu(m_e^2) = w_e^1$ and $\mu(m_e^1) = w_e^2$. Moreover, denote $M_E^S = \{m_e^i \in M_E : e \in S\}$. If there existed $\{u, v\} \in S$ such that $u \notin U$, then $(m_{\{u,v\}}^1, w_u^1)$ would have formed a blocking pair, which contradicts the fact that μ is a stable matching. Thus, we have that the set of endpoints of the edges in S is a subset of U.

We claim that $|U| = k$ and that U is the vertex-set of a clique in G, which would imply that $\mathcal{I}$ is a Yes-instance. Since we have argued that the set of endpoints of the edges in S is a subset of U, it is sufficient to show that $|U| \leq k$ and $|S| \geq \dfrac{k(k-1)}{2}$ (note that $|S| \geq \dfrac{k(k-1)}{2}$ implies that $|U| \geq k$), as this would imply that U is indeed the vertex-set of a clique on k vertices in G. First, since $\mathsf{balance}(\mu) \leq \widehat{k}$, we have that $\displaystyle\sum_{(m,w) \in \mu} p_m(w) \leq |M| + \delta + 6(k + \dfrac{k(k-1)}{2})$. Now, note that

$$
\begin{aligned}
\sum_{(m,w) \in \mu} p_m(w) &= \sum_{m \in M_V^U} p_m(\mu(m)) + \sum_{m \in M_V \setminus M_V^U} p_m(\mu(m)) + \sum_{m \in M_E^S} p_m(\mu(m)) \\
&\quad + \sum_{m \in M_E \setminus M_E^S} p_m(\mu(m)) + \sum_{m \in \widetilde{M}} p_m(\mu(m)) + p_{m^*}(\mu(m^*)) \\
&= 4|M_V^U| + |M_V \setminus M_V^U| + 4|M_E^S| + |M_E \setminus M_E^S| + |\widetilde{M}| + \delta + 1 \\
&= |M| + \delta + 6(|U| + |S|).
\end{aligned}
$$

Thus, we deduce that $|U| + |S| \leq k + \dfrac{k(k-1)}{2}$. Now, observe that since $\mathsf{balance}(\mu) \leq \widehat{k}$, we also have that $\displaystyle\sum_{(m,w) \in \mu} p_w(m) \leq |M| + \delta + 6(k + \dfrac{k(k-1)}{2})$. Here, on the one hand we note that

$$
\begin{aligned}
\sum_{(m,w) \in \mu} p_w(m) &= \sum_{m \in M_V^U} p_{\mu(m)}(m) + \sum_{m \in M_V \setminus M_V^U} p_{\mu(m)}(m) + \sum_{m \in M_E^S} p_{\mu(m)}(m) \\
&\quad + \sum_{m \in M_E \setminus M_E^S} p_{\mu(m)}(m) + \sum_{m \in \widetilde{M}} p_{\mu(m)}(m) + p_{\mu(m^*)}(m^*) \\
&= |M_V^U| + |M_V \setminus M_V^U|(|E| + 2) + |M_E^S| + |M_E \setminus M_E^S|(|V||E| + 2) + |\widetilde{M}| + 1 \\
&= |M| + 2(|V| - |U|)(|E| + 1) + 2(|E| - |S|)(|V||E| + 1) \\
&= |M| + 2(|V| + |E| + |V||E| + |V||E|^2) - 2|U|(|E| + 1) - 2|S|(|V||E| + 1).
\end{aligned}
$$

On the other hand, we note that

$$
\begin{aligned}
\widehat{k} &= |M| + \delta + 6(k + \frac{k(k-1)}{2}) \\
&= |M| + 2(|V| + |E| + |V||E| + |V||E|^2) - k(4 + 4k + 2|E| + (k-1)|V||E|) + 6(k + \frac{k(k-1)}{2}) \\
&= |M| + 2(|V| + |E| + |V||E| + |V||E|^2) - 2k(|E| + 1) - k(k-1)(|V||E| + 1).
\end{aligned}
$$

Thus, we have that

$$
|U|(|E| + 1) + |S|(|V||E| + 1) \geq k(|E| + 1) + \frac{k(k-1)}{2}(|V||E| + 1)
$$

Recall that we have also shown that $|U| + |S| \leq k + \frac{k(k-1)}{2}$. Thus, since $|U| \leq k + \frac{k(k-1)}{2} < |V|$, to satisfy the above equation it must hold that $|S| \geq \frac{k(k-1)}{2}$. Since $|U| + |S| \leq k + \frac{k(k-1)}{2}$, we deduce that $|U| \leq k$. This, as we have argued earlier, finished the proof. $\square$

This concludes the proof of Theorem 5.1.

References

[1] J. Chen, D. Hermelin, M. Sorge, and H. Yedidsion, *How hard is it to satisfy (almost) all roommates?*, arXiv preprint arXiv:1707.04316, (2017).

[2] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh, *Parameterized Algorithms*, Springer, 2015.

[3] M. Cygan, M. Pilipczuk, M. Pilipczuk, and J. O. Wojtaszczyk, *On multiway cut parameterized above lower bounds*, TOCT, 5 (2013), pp. 3:1–3:11.

[4] R. G. Downey and M. R. Fellows, *Fixed-parameter tractability and completeness II: On completeness for W[1]*, Theoretical Computer Science, 141 (1995), pp. 109–131.

[5] ———, *Fundamentals of parameterized complexity*, Springer, 2013.

[6] T. Feder, *Stable networks and product graphs*, PhD thesis, Stanford University, 1990.

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

[8] ———, *College admissions and the stability of marriage*, American Mathematical Monthly, 69 (1962), pp. 9–15.

[9] D. Gale and M. Sotomayor, *Some remarks on the stable matching problem*, Discrete Applied Mathematics, 11 (1985), pp. 223–232.

[10] S. Garg and G. Philip, *Raising the bar for vertex cover: Fixed-parameter tractability above A higher guarantee*, in Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016, 2016, pp. 1152–1166.

[11] D. Gusfield and R. W. Irving, *The Stable marriage problem - structure and algorithms*, Foundations of computing series, MIT Press, 1989.

[12] R. W. Irving, P. Leather, and D. Gusfield, *An efficient algorithm for the "optimal" stable marriage*, Journal of ACM, 34 (1987), pp. 532–543.

[13] A. Kato, *Complexity of the sex-equal stable marriage problem*, Japan Journal of Industrial and Applied Mathematics, 10 (1993), p. 1.

[14] D. E. Knuth, *Stable marriage and its relation to other combinatorial problems : an introduction to the mathematical analysis of algorithms*, CRM proceedings & lecture notes, Providence, R.I. American Mathematical Society, 1997.

[15] D. Lokshtanov, N. S. Narayanaswamy, V. Raman, M. S. Ramanujan, and S. Saurabh, *Faster parameterized algorithms using linear programming*, ACM Trans. Algorithms, 11 (2014), pp. 15:1–15:31.

[16] D. F. Manlove, *Algorithmics of Matching Under Preferences*, vol. 2 of Series on Theoretical Computer Science, WorldScientific, 2013.

[17] E. McDermid, in Personal communications between McDermid and Manlove, 2010.

[18] E. McDermid and R. W. Irving, *Sex-equal stable matchings: Complexity and exact algorithms*, Algorithmica, 68 (2014), pp. 545–570.

[19] G. O'Malley, *Algorithmic aspects of stable matching problems*, PhD thesis, University of Glasgow, 2007.

[20] V. Raman, M. S. Ramanujan, and S. Saurabh, *Paths, flowers and vertex cover*, in Algorithms - ESA 2011 - 19th Annual European Symposium, Saarbrücken, Germany, September 5-9, 2011. Proceedings, 2011, pp. 382–393.