Coalition Manipulations in the Gale-Shapley Algorithm

Yuan Deng, Duke University
Weiran Shen, Tsinghua University
Pingzhong Tang, Tsinghua University

It is well-known that the Gale-Shapley algorithm is not truthful for all agents. Previous studies in this category concentrate on manipulations using incomplete preference lists by a single woman and by the set of all women. Little is known about manipulations by a subset of women or other types of manipulations, such as permutation of complete preference lists.

In this paper, we consider manipulations by any subset of women with arbitrary preferences (either incomplete or complete). For the setting where agents can report an incomplete preference list (aka. general manipulations), we show that a strong Nash equilibrium of the induced manipulation game always exists among the manipulators and the equilibrium outcome is unique and Pareto-dominant. In addition, the set of matchings achievable by manipulations has a lattice structure.

For the setting where agents can only report complete preference lists (aka. permutation manipulations), we give answers to Gusfield and Irving's open question on what matchings can be achieved in the induced manipulation games. We first construct a counter-example to show that a Pareto-dominant outcome may not exist. Then we present a polynomial-time algorithm to find a Pareto-optimal strategy profile for the induced manipulation game. Furthermore, we show that Pareto-optimality is equivalent to super-strong Nash equilibrium outcomes and all such matchings can be found by our algorithm. The results for the second part are enabled by connecting this problem to the stable roommate problem and using techniques there to analyze a graph called suitor graph. We also introduce several new concepts, such as maximum rotation and principle set, and develop a series of original techniques.

Even though all these results may suggest that the Gale-Shapley algorithm is vulnerable to coalition manipulations, we do, however, prove a hardness result in the end, saying that it is NP-complete to find a manipulation that induces a matching strictly better off for all manipulators.

Additional Key Words and Phrases: Stable matching, algorithm, complexity, Gale-Shapley algorithm, coalition manipulation

1. INTRODUCTION

Stability has been a central concept in economic design, ever since the seminal work by Gale and Shapley [1962]. Over the years, intensive research has been done in the literature of stable matching. A variety of applications of this problem have also been developed, ranging from college admissions and school matchings [Abdulkadiroğlu et al. 2005; Abdulkadiroglu and Sönmez 2003; Gale and Shapley 1962] to centralized kidney exchange programs [Abraham et al. 2007; Roth et al. 2004, 2005] and hospitals-residents matchings [Irving and Manlove 2009; Irving et al. 2000; Roth 1996].

In the standard stable matching model, there is a set of men and a set of women. Each agent has a preference list over a subset of the opposite sex. A matching between men and women is stable if no pair of agents prefer to match with each other than their designated partner. Gale and Shapley [1962] put forward an algorithm, aka. the Gale-Shapley algorithm, that computes a stable matching in $O(n^2)$ time. The algorithm (men-proposing version) proceeds in multiple rounds. At each round, each man

This work was supported by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the Natural Science Foundation of China Grant 61033001, 61361136003, 61303077, 61561146398, a Tsinghua Initiative Scientific Research Grant and a China Youth 1000-talent program.

Authors addresses: Yuan Deng, Computer Science Department, Duke University; email: ericdy@cs.duke.edu; Weiran Shen, IIIS, Tsinghua University; email: emersonswr@gmail.com;Pingzhong Tang, IIIS, Tsinghua University; email: kenshinping@gmail.com;

proposes to his favorite woman that has not rejected him yet; and each woman keeps her favorite proposal, if any, and rejects all others. The algorithm iterates until no further proposal can be made.

The algorithm enjoys many desirable properties. It is well-known that the matching returned by the algorithm is preferred by every man to any other stable matching, hence called the M-optimal (for men-optimal) matching. It is also known that all stable matchings form a lattice defined by such a preference relation and the M-optimal matching is the greatest element in the lattice [Knuth 1976]. Furthermore, men and women have strictly opposite preferences over two stable matchings: every man prefers stable matching μ_1 to stable matching μ_2 if and only if every woman prefers μ_2 to μ_1. As a result, the M-optimal matching is the W-pessimal (for women-pessimal) matching [McVitie and Wilson 1971]. The smallest element in the lattice, the W-optimal (M-pessimal) matching, can be obtained by swapping the roles of men and women.

A concern with the Gale-Shapley algorithm is its non-truthfulness. While it is known that the algorithm is group strategy-proof[1] for all men [Dubins and Freedman 1981], it is not truthful for women. In fact, Roth [1982] shows that there is no stable matching algorithm that is strategyproof for all agents.

1.1. Related work

The above impossibility theorem by Roth initiates an interesting literature of finding manipulations for women by fixing men preferences in the Gale-Shapley algorithm. Knuth et al. [1990] show that any woman may have at least $\left(\frac{1}{2} - \epsilon\right) \ln n$ and at most $(1 + \epsilon) \ln n$ different partners in all possible stable matchings, where n is the number of men and ϵ is a positive constant. Gale and Sotomayor [1985] show that it is possible for all women to strategically truncate their preference lists so that each of them is matched with their partner in the W-optimal matching, and Teo et al. [2001] provide a polynomial time algorithm to find the optimal single-agent truncation manipulation.

Jaramillo et al. [2014] study dropping strategies in a many-to-many setting where agents can use blacklists but no shuffling is allowed. They show that dropping strategies are exhaustive in this setting, i.e., any stable matching can be replicated or improved using some dropping strategies. Gonczarowski [2014] also looks into general manipulations where the agents use dropping strategies. He focuses on manipulations by all women and give a tight upper bound on the number of men that must be deleted to reach the W-optimal matching in the worst case.

Teo et al. [2001] study permutation manipulations, where a woman can report any permutation of her true preference list[2]. They give an efficient algorithm to compute the best manipulation for a single manipulator. Aziz et al. [2015] also study permutation manipulations in a many-to-one setting, but focus on a single manipulator with quota more than one. Huang [2010] considers the classified stable matching problem in a many-to-one setting where each man has his own classifications and has quota limits on each class. Pini et al. [2009] create a stable matching mechanism and show that for a single agent, it is computationally hard to manipulate the matching result. All the results, except for the last, do not apply to cases where a coalition of women jointly manipulate.

[1] Precisely, *group strategy-proof* means no coalition manipulation can make all men in the coalition strictly better off, in this context. If consider the case where no man is worse off and at least one man is strictly better off, the Gale-Shapley algorithm is not group strategy-proof [Huang 2006].

[2] This type of manipulations is common in the voting domain, e.g. Gibbard [1973].

1.2. Our contributions

In this paper, we study the game where a coalition of women can manipulate the Gale-Shapley algorithm. The manipulators can use either general manipulation or permutation manipulations.

We first analyze the setting of general manipulations, where agents (women) can report a preference list over any *subset* of men. We show that a strong Nash equilibrium (i.e,. no subset of manipulators can deviate and get strictly better off) always exists for any subset of women.

THEOREM 1.1. *In general manipulations, it is a strong Nash equilibrium that each manipulator removes every man below her W-optimal partner on her list. Furthermore, in the induced matching, all manipulators can be matched to their W-optimal partners.*

This result generalizes the result by Gale and Sotomayor [1985], which considers manipulations by the set of all women, and Teo et al. [2001], which provides an efficient algorithm to compute the optimal single-agent general manipulation. Moreover, the equilibrium outcome is unique and Pareto-dominant for all manipulators, i.e., all manipulators reach a consensus on a single manipulation profile. Furthermore, the set of all stable matchings attainable from general manipulations forms a join-semilattice (Theorem 3.7).

For the more challenging setting of permutation manipulations where agents can permute their preference lists, none of the nice properties in the general manipulation setting continues to hold. We first give an example to show that a coalition of women could get worse off by manipulating jointly than each performing a single-agent manipulation (see Table I for details). As a result, unlike the general manipulation setting, a unique Pareto-dominant outcome for all manipulators may not exist.

THEOREM 1.2. *In permutation manipulations, there exists a polynomial-time algorithm to find a manipulation such that the induced matching is Pareto-optimal and stable with respect to true preference lists.*

In fact, all Pareto-optimal matchings can be found by the algorithm. The algorithm iteratively improves the matching and checks if there exists a strategy profile for the manipulators that induces such a matching. We thus give an algorithmic characterization of the open problem raised by Gusfield and Irving [1989] on what can be produced by permutation manipulations. This problem is studied and reemphasized as an open problem in [Kobayashi and Matsui 2010] and [Sukegawa and Yamamoto 2012].

Our algorithm is enabled by connecting stable matching problem to stable roommate problem [Irving 1985], and by extensively using a structure called *rotation* defined in the stable roommate problem. Using this connection, we can apply an important tool called *suitor graph* [Kobayashi and Matsui 2010]. The suitor graph is constructed using a desired matching and the preference profile of all truthful agents. It can help construct the manipulators' preference profile that would yield the desired matching. To obtain our results, we first construct the suitor graph according to the W-pessimal matching. Then we eliminate rotations and modify the suitor graph accordingly. We can decide whether the current matching corresponds to feasible manipulations by checking certain connectivity properties of the suitor graph. However, some rotations become exposed only after others are eliminated and thus only closed sets of rotations can be validly eliminated. We then introduce the concept of *maximal rotations* and prove that only closed sets whose maximal rotations contain manipulators can induce feasible manipulations. However, we cannot afford to enumerate all closed sets since the number grows exponentially. To tackle this, we further introduce the concept of *principle sets*, whose number is polynomial with respect to the number of men and women, and show that any feasible closed sets can be found if we eliminate feasible principle sets iteratively. A Pareto-optimal strategy profile is found when no such principle set can be eliminated.

We also prove another conceptually interesting result that only minor modifications are needed for the manipulators to perform a Pareto-optimal manipulation (Theorem 4.20), another evidence of the vulnerability of the algorithm. Furthermore, we show that Pareto-optimality is equivalent to super-strong Nash equilibrium, i.e., no subset of manipulators can deviate and get weakly better off and at least one gets strictly better off (Theorem 4.21).

Our results on permutation manipulations can be seen as a novel generalization of the main result of Teo et al. [2001], who present an algorithm to compute strategies for one manipulator by exhaustively searching for possible partners. However, directly extending their results to coalition manipulations suffers from unacceptable time complexity, since we need to enumerate all possible partner combinations. Our techniques differ from theirs substantially and can provide different insights of the problem.

Although these results may suggest that the Gale-Shapley algorithm is vulnerable to coalition manipulations, the following hardness result shows it is NP-complete to find a manipulation such that the induced matching is strictly better off for all manipulators.

THEOREM 1.3. *In permutation manipulations, it is NP-complete to find a manipulation such that the induced matching is strictly better off for all manipulators and stable with respect to true preference lists.*

In other words, if there is a small manipulation cost for the agents, then: (1) any agent is reluctant to participate in a manipulation that is not strictly better for her (2) any manipulation that is strictly better off is unlikely to be found.

2. PRELIMINARIES

We consider a stable matching model with a set of men M and a set of women W and assume $|M| = |W|$. The preference list of a man m, denoted by $P(m)$, in a preference profile P is a strict total order $\succ_m^P$ over a subset of W. Let $w_1 \succ_m^P w_2$ denote that m prefers w_1 to w_2 in profile P. Similarly, the preference list of a woman w is a strict total order over a subset of M. For simplicity, we use $\succ_w$ to denote the true preference profile when it is clear from the context. Denote the preference profile of M and W by $P(M)$ and $P(W)$, respectively. A matching is a function $\mu : M \cup W \to M \cup W$. We write $\mu(m) = w$ if a man m is matched to a woman w, or $\mu(m) = m$ if he is unmatched. Similarly, $\mu(w) = m$ if w is matched to m and $\mu(w) = w$ if unmatched. Moreover, for two matchings μ_1 and μ_2, if for all $w \in W$, $\mu_1(w) \succeq_w \mu_2(w)$, we say $\mu_1 \succeq_W \mu_2$.

A matching is *individually rational* if no one is matched to someone who is not in his or her preference list. If in a matching μ, a man m and a woman w are not matched together, yet prefer each other to their partners in μ, then (m, w) is called a *blocking pair*. A matching is *stable* if it is individually rational and contains no blocking pair.

The Gale-Shapley algorithm is not truthful for women [Dubins and Freedman 1981]. Given a set of women manipulators, the algorithm can be thought of as a game (henceforth, the *manipulation game*), between them. Let $L \subseteq W$ be the set of manipulators and $N = W \setminus L$ be the set of non-manipulators.

Definition 2.1 (*Manipulation game*). Given a true preference profile P, a manipulation game is a tuple (L, A^L), where:

(1) $L \subseteq W$ is the set of manipulators;
(2) $A^L = \prod_{i \in L} A_i$ is the set of all possible reported preference profiles.

The outcome of the manipulation game (also called induced matching in this paper) is the matching resulted from the Gale-Shapley algorithm with respect to the reported preference profiles. A manipulator's preference in this game is naturally her true preference in P.

Remark 2.2. In a manipulation game, only manipulators L are considered as players, i.e., the reported preference profiles for all men and non-manipulators are their true preference profiles. The set of all possible preferences A_i depends on different manipulation types and we only consider the case where all manipulators use the same type of manipulations.

We now define three types of manipulations [Gonczarowski 2014; Kobayashi and Matsui 2010; Roth and Sotomayor 1992], which determines the elements in A^L.

Definition 2.3 (*General manipulation*). Let $\mathbb{O}_b$ be the set of strict total orders over any subset of M. In general manipulations, $A_i = \mathbb{O}_b, \forall i \in L$.

Definition 2.4 (*Truncation manipulation*). Let $(m_1, m_2, \ldots, m_k)$ be a woman i's true preference list. In truncation manipulations, $A_i = \{(m_1, m_2, \ldots, m_j) \mid \forall j \leqslant k\}$, $\forall i \in L$.

Definition 2.5 (*Permutation manipulation*). Let $\mathbb{O}_p$ be the set of strict total orders over M. In permutation manipulations, $A_i = \mathbb{O}_p$, $\forall i \in L$.

Clearly, the truncation manipulation is a special case of the general manipulation. In permutation manipulations, each woman takes interest in all men and a manipulation is simply a total order on M.

Let $P(L)$ be the preference profile of all manipulators. We slightly abuse notations and write $P(L) = \bigcup_{l \in L} P(l)$, where $P(l)$ is the preference list reported by l. Similarly, denote the preference profiles for men and for non-manipulators by $P(M)$ and $P(N)$. Thus the overall preference profile is $P = (P(M), P(N), P(L))$. Denote by $S(P(M), P(W))$ the set of all stable matchings under profile $(P(M), P(W))$.

In this paper, our results are based on the feasibility assumption of a manipulation, according to a series of literature [Gale and Sotomayor 1985; Roth and Vande Vate 1991; Roth 2002].

ASSUMPTION 2.6 (FEASIBILITY ASSUMPTION). *A manipulation is feasible if the induced matching is stable with respect to the* true *preference profile.*

We call a matching *feasible* if it is induced by a feasible manipulation. The assumption is worth some explanations. As Roth [2002] and Roth and Vande Vate [1991] suggest, stability is of great importance for a successful clearinghouse. Empirical evidence shows that most stable mechanisms have succeeded in practice while almost all unstable ones have failed. If the induced matchings were unstable, a manipulation is no longer a Nash equilibrium, thus agents are unlikely to follow and the manipulations fall apart [Gale and Sotomayor 1985]. Such unpredictability makes the unstable matchings less desirable. This is equivalent to what the assumption states: unstable matchings yield low payoffs for the agents. The following solution concepts are all based on the feasibility assumption.

Definition 2.7 (*Nash equilibrium*). A preference profile $P(L) = \bigcup_{l \in L} P(l)$ of a manipulation game is a Nash equilibrium if $\forall l \in L$, $P(l)$ is a best response to $P(L) \setminus \{P(l)\}$.

In other words, in a Nash equilibrium, l cannot be matched with a better partner in any stable matching she can manipulate to.

Definition 2.8 (*Strong Nash equilibrium & Super-strong Nash equilibrium*). A Nash equilibrium is *strong*, if no subset of manipulators can jointly manipulate to a matching that is strictly better off for all of them. A Nash equilibrium is *super-strong*, if no subset of manipulators can jointly manipulate to a matching that is weakly better off for all and strictly better off for at least one of them.

Definition 2.9 (*Pareto-optimal matching*). A matching μ is Pareto-optimal if there is no feasible matching in which all women (not necessarily a manipulator) are weakly better off than in μ and at least one woman is strictly better off.

The above definition can be thought of as the standard Pareto-optimality restricted to feasible matchings. Let $S_A(P(M), P(W))$ denote the set of all feasible matchings. We sometimes write S_A for short when $(P(M), P(W))$ is clear from the context. Finally, we say a strategy profile is Pareto-optimal when its induced matching is Pareto-optimal.

3. GENERAL MANIPULATIONS

Gale and Sotomayor [1985] prove that a strong Nash equilibrium always exists if all women are manipulators and use truncation manipulations. They construct explicitly such a strong equilibrium by letting each woman use a truncation manipulation that removes all men ranked below her W-optimal partner. Ma [2010] also studies truncation manipulations in the same setting and shows that there is only one Nash equilibrium. In addition, the equilibrium profile admits a unique stable matching, namely, the W-optimal matching. Moreover, Teo et al. [2001] provide a polynomial time algorithm to find the optimal single-agent manipulation. In this section, we extend these results to coalition manipulations and consider any subset $L \subseteq W$ as manipulators.

LEMMA 3.1. *Let $P = (P(M), P(W))$ be the true preference profile for all agents. Every matching in $S_A(P)$ induced by a feasible general manipulation can be induced by a feasible truncation manipulation.*

Remark 3.2. Note that this result is different from the exhaustiveness result in [Jaramillo et al. 2014], since exhaustiveness only requires that $\forall \mu \in S_A(P)$, there exists a truncation manipulation such that the induced matching is *weakly preferred* by the manipulators.

According to Lemma 3.1, it is therefore without loss of generality to focus on truncation manipulations. In the remainder of this section, unless explicitly specified, we say a partner or a matching is W-optimal or W-pessimal for a woman if it is so under the true preference profile.

3.1. Super-strong Nash equilibria

The following theorem states that any unmatched woman in a stable matching remains unmatched in all stable matchings.

THEOREM 3.3 (ROTH [1986]). *Given $P(M)$ and $P(W)$, the set of unmatched agents is the same among all stable matchings.*

Recall that a feasible manipulation must induce a stable matching under true preferences. Therefore, any unmatched woman in the W-optimal matching has no incentive to misreport since she will always be unmatched. Thus, we only need to consider the case where no manipulator is unmatched in the W-optimal matching.

THEOREM 3.4 (THEOREM 1.1). *In truncation manipulations, it is a super-strong Nash equilibrium that each manipulator removes every man below her W-optimal partner on her list. Furthermore, in the induced matching, all manipulators can be matched to their W-optimal partners.*

This result generalizes the result by Gale and Sotomayor [1985], which only considers manipulations by the set of all women. If the set of manipulators contains only one woman, the problem becomes a single-agent manipulation and Theorem 3.4 can also be applied. Thus, in coalition manipulations, every manipulator is matched with the same man as in her best single-agent manipulation.

3.2. Lattice structure

Definition 3.5. Given two matchings μ and μ', define $\mu_\vee = \mu \vee \mu'$ to be the matching that matches each man to his more preferred partner and each woman to her less preferred partner in μ and μ'. Similarly, we can define $\mu_\wedge = \mu \wedge \mu'$, which matches each man to his less preferred partner and each woman to her more preferred partner.

The following theorem states that $\mu_\vee$ and $\mu_\wedge$ are not only well-defined matchings, but also essential to the lattice structure of the set of all stable matchings.

THEOREM 3.6 (CONWAY'S LATTICE THEOREM; [KNUTH 1976]). *When all preferences are strict, if μ and μ' are stable matchings under preference profile P, then the matching $\mu_\vee = \mu \vee \mu'$ and $\mu_\wedge = \mu \wedge \mu'$ are both matchings. Furthermore, they are both stable under P.*

Therefore, the set of all stable matchings is a lattice with $\succeq_M$ and $\succeq_W$. Let μ^L be a partial matching obtained by restricting the corresponding full matching μ to the set of manipulators and we call μ an *extension* of μ^L. Let S_A^L be the set of partial matchings obtained by restricting all matchings in S_A to the set of manipulators L.

THEOREM 3.7. *Given the true preference profiles $(P(M), P(W))$ and the set of manipulators L, then the set of matchings that can be induced by feasible general manipulations, $S_A(P(M), P(W))$, is a join-semilattice*[3]*, and the set of partial matchings $S_A^L(P(M), P(W))$ is a lattice.*

Since a finite join-semilattice has a greatest element, if there exist two distinct matchings resulted from super-strong Nash equilibria, at least one matching can be improved. Thus, the matching induced from the super-strong Nash equilibria is *unique* and *Pareto-dominant*.

3.3. Relaxing the feasibility assumption

Recall that we consider feasible manipulations and if this restriction is relaxed, the strategy profile constructed in Theorem 3.4 is still a strong Nash equilibrium.

THEOREM 3.8. *In general manipulations without the feasibility assumption, it is a strong Nash equilibrium that each manipulator removes every man below her W-optimal partner on her list. Furthermore, in the induced matching, all manipulators can be matched to their W-optimal partners.*

However, without the feasibility assumption, it is no longer a super-strong Nash equilibrium. A counter-example is provided in Appendix C.1.

4. PERMUTATION MANIPULATIONS

In this section, we analyze the open problem raised by Gusfield and Irving [1989] on what can be achieved by permutation manipulations. We are interested in algorithmic characterizations of solution concepts such as Pareto-optimal strategy profiles and super-strong Nash equilibria. Formally, we have the following results.

[3] A join-semilattice is a partially ordered set where every two elements have a unique join (or supremum).

THEOREM 4.1 (FORMAL VERSION OF THEOREM 1.2). *There exists a polynomial-time algorithm in the number of agents, such that given any complete preference profile P and any subset $L \subseteq W$ as manipulators, the algorithm computes a feasible permutation manipulation, i.e., a strategy profile $P'(L)$, for L and the induced matching μ':*

(*1*) *if L reports $P'(L)$, the induced matching μ' is Pareto-optimal;*
(*2*) *for each $w \in L$, $P'(w)$ is modified from $P(w)$ by moving at most one man to some higher ranking.*

Moreover, the algorithm can be adapted to find the set of all Pareto-optimal matchings.

Remark 4.2. It is weakly better off for all manipulators to follow the strategy $P'(L)$ rather than P, since μ' is stable under P, which is preferred by each manipulator to the W-pessimal matching under P.

THEOREM 4.3. *Under the feasibility assumption, the set of all Pareto-optimal matchings is the same as the set of super-strong Nash equilibrium outcomes.*

The two theorems together indicate that all super-strong Nash equilibria can be found by our algorithm. Recall that in permutation manipulations, both men and women have complete preference lists and each manipulator is only allowed to permute her true preference list. Previous results about general manipulation do not carry over here mainly due to the challenge that joint manipulations may result in worse matchings. Consider an example with 4 men and 4 women (see Table I).

Table I. Example of non-cooperativeness

Men's preference lists					Women's preference lists				
m_1	w_1	w_4	w_2	w_3	w_1	m_3	m_2	m_1	m_4
m_2	w_1	w_3	w_2	w_4	w_2	m_1	m_4	m_3	m_2
m_3	w_2	w_3	w_1	w_4	w_3	m_2	m_3	m_1	m_4
m_4	w_2	w_4	w_1	w_3	w_4	m_4	m_1	m_3	m_2

$\{(m_1, w_4), (m_2, w_1), (m_3, w_3), (m_4, w_2)\}$ is the M-optimal matching. Suppose the set of manipulators is $L = \{w_1, w_2\}$ and consider individual manipulations by w_1 and w_2.

(1) w_1 exchanges m_1 and m_2 and get $\{(m_1, w_4), (m_2, w_3), (m_3, w_1), (m_4, w_2)\}$;
(2) w_2 exchanges m_3 and m_4 and get $\{(m_1, w_2), (m_2, w_1), (m_3, w_3), (m_4, w_4)\}$;

In both cases, w_1 and w_2 can manipulate to get their W-optimal partner. However, if they try to cooperate, it is surprising that they both get worse off than the matching corresponding to their true preference lists. In fact, the only way for them to manipulate together is that each manipulator performs the operation mentioned above and the induced matching is $(m_1, w_1), (m_2, w_3), (m_3, w_2), (m_4, w_4)$.

The above example shows that conflict of interest exists among different manipulators. To analyze the problem, we borrow two structures, rotations [Irving 1985] and suitor graphs [Kobayashi and Matsui 2009], from the literature. We further develop several new structures such as maximal rotations and principle sets to derive connections between suitor graphs and feasible manipulations.

The remainder of this section is organized as follows: Section 4.1 and Section 4.2 briefly introduce the structures of rotations and suitor graphs. Section 4.3 and Section 4.4 provides an algorithmic characterization of Pareto-optimal strategy profiles. Section 4.5 characterizes the equivalence between Pareto-optimal matchings and super-strong Nash equilibrium outcomes. Section 4.6 discusses variations of the problem and provides hardness results for a decision problem and a counting problem.

4.1. The stable roommate problem and rotations

The stable roommate problem is a natural generalization of the stable marriage problem. In the stable roommate problem, each agent has a preference list over all other agents and can be matched to any other agent. We abbreviate the two problems to SM (stable marriage) and SR (stable roommate) respectively.

Irving [1985] designs an efficient algorithm for solving the SR problem. The algorithm consists of two phases. In phase 1, the algorithm runs just like the Gale-Shapley algorithm. Each agent proposes to other agents according to his preference list. When a_i proposes to a_j, a_j accepts a_i if he does not hold any proposal or a_i is better than his current mate, and rejects a_i otherwise. If a_i is rejected, then remove them from each other's preference list. It is shown that if a_j accepts a_i, then a_j can not be matched with anyone ranked below a_i in a_j's preference list. As a result, if a_i is accepted, for each a_k ranked below a_i in a_j's preference list, we can remove a_k and a_j from each other's preference list. Phase 1 terminates until each agent holds a proposal. Each agent's list at the end of phase 1 is called a reduced list. The set of all reduced lists is called a reduced table. At the end of phase 1, a_i is in a_j's reduced list if and only if a_j is in a_i's. In phase 2, the algorithm solves the problem by eliminating a series of rotations.

Definition 4.4 (*Rotation*). A rotation is a sequence of agents $R = (a_1, a_2, \dots, a_r)$ where the first entry of a_{i+1}'s reduced list is the second in a_i's reduced list, for all $1 \le i \le r$, and $i + 1$ is taken modulo r.

The elimination of a rotation R is to force an agent a_i of R to reject his current proposer. Then we can run the phase 1 algorithm since a_i no longer holds any proposal. the rotation R is eliminated until the phase 1 algorithm terminates again. After each elimination, the reduced table is updated and new rotations may appear. We call a rotation *exposed* if it is in the current reduced list. The phase 2 algorithm terminates until no rotations are exposed. We refer readers to the original paper for more details of the algorithm [Irving 1985].

Gusfield [1988] explores the structure of the solutions of the SR problem. He discovered that some rotations are singleton while others have a *dual rotation*. A dual rotation of a rotation $R = (a_1, a_2, \dots, a_r)$ is also a rotation $R^d = (a_1^d, a_2^d, \dots, a_r^d)$ where R and R^d have the same length and a_i^d is the second entry in a_i's reduced list. Gusfield [1988] proves that in order to find a solution, all singleton rotations and exactly one rotation of each dual pair must be eliminated. The solution is determined by the set of eliminated rotations but not the order of elimination. Furthermore, he showed that every solution corresponds to a certain set of rotations and there exists an order of elimination that produces it.

We also use $R = (\mathcal{A}, \mathcal{F}, \mathcal{S})$ to represent a rotation, where $\mathcal{A}$ is the sequence of agents contained in R and $\mathcal{F}$ and $\mathcal{S}$ are the corresponding sequences of the first and the second entry in $\mathcal{A}$'s reduced lists. Therefore, $\mathcal{F}_{i+1} = \mathcal{S}_i$ by the definition of rotations, and $R^d = (\mathcal{S}, \mathcal{A}, \mathcal{A}^r)$, where $\mathcal{A}^r$ is the sequence $\mathcal{A}$ with each agent shifted left by one position.

A SM problem can be easily converted to a SR problem by adding all agents of the same sex to the end of each agent's preference list. One can easily check that the converted problem has exactly the same set of solutions as the original problem. We study the rotations of the converted problem and figure out how to *eliminate* rotations in the Gale-Shapley algorithm. Although rotations can be defined directly in the SM problem, known as *improvement cycle* and useful in the algorithm that converts M-optimal matching to W-optimal matching [Ashlagi et al. 2013; Gonczarowski 2014; Immorlica and Mahdian 2005] , we analyze from a different point of view, which can provide more insight into the structure of the problem. Note that the converted SR problem is only used to analyze the structure of the rotations, we still stick to the Gale-Shapley algorithm to solve the original SM problem.

LEMMA 4.5. *Any SM problem can be converted to a SR problem using the method above. In the corresponding SR problem, each rotation contains only agents of the same sex and each rotation has a dual rotation.*

We call a rotation containing only men (women) a M-rotation (W-rotation). Moreover, an M-rotation's dual is a W-rotation and vice versa. Therefore, all pairs of rotations contain an M-rotation and a W-rotation. To generate a certain solution to the converted SR problem, we need to select either an M-rotation or a W-rotation to eliminate in each dual pair. Thus, each solution can be represented using a set of M-rotations, and given a set of M-rotations, we can construct its corresponding reduced table, and the corresponding matching just matches each man to the first woman in his reduced list (or matches each woman to the last man in her reduced list) after eliminating the given set of M-rotations.

We use *rotations* to represent M-rotations from now on since only M-rotations need to be considered. A man m is in a rotation $R = (\mathcal{M}, \mathcal{W}, \mathcal{W}^r)$ if m is in the sequence $\mathcal{M}$ and a woman w is in R if w is in $\mathcal{W}$. Also, an agent is in a set of rotations if this agent is in any rotation of the set. We may sometimes use m_i and w_i to mean the i-th agent in $\mathcal{M}$ and $\mathcal{W}$ when the order is important. Moreover, we say a rotation $R = (\mathcal{M}, \mathcal{W}, \mathcal{W}^r)$ moves m_i from w_i to w_{i+1} and moves w_i from m_i to m_{i-1} since after eliminating the rotation, the corresponding matching matches m_i and w_{i+1} together.

It is known that the order of proposals and rejections does not affect the induced matching. As a result, in order to eliminate a rotation R and compute the induced matching, we can pick an arbitrary woman in R and let her reject her current proposer. Then we run the Gale-Shapley algorithm again until it terminates.

4.2. Suitor graph

Suitor graph is another important structure for our analysis. It is proposed by Kobayashi and Matsui [2010] when considering the problem that given a preference profile for all truthful agents $P(M)$ and $P(N)$, is there a profile $P(L)$ for the manipulators such that the M-optimal matching of the combined preference profile is a certain matching μ? The detailed definition of suitor graph is as follows:

Definition 4.6 (*Suitor graph; Kobayashi and Matsui [2010]*). Given a matching μ, a preference profile for all men $P(M)$ and a preference profile for all non-manipulators $P(N)$, the corresponding suitor graph $G(P(M), P(N), \mu)$ is a digraph (V, E). Then, $G(P(M), P(N), \mu)$ can be constructed using the following steps:

(1) $V = M \cup W \cup s$, where s is a virtual vertex;
(2) $\forall w \in W$, add edges $(w, \mu(w))$ and $(\mu(w), w)$;
(3) $\forall w \in W$, let $\delta(w) = \{m \mid w \succ_m \mu(m)\}$;
(4) $\forall w \in L$ and for each m in $\delta(w)$, add edges (m, w);
(5) $\forall w \in N$, if $\delta(w)$ is nonempty, add the edge (m, w), where m is w's favorite in $\delta(m)$;
(6) $\forall w \in W$, if $\delta(w) = \emptyset$, add an edge (s, w) to the graph;

Kobayashi and Matsui [2010] also give a characterization of the existence of such profiles and a $O(n^2)$ time algorithm can be found directly from their constructive proof.

THEOREM 4.7 (KOBAYASHI AND MATSUI [2010]). *Given a matching μ, a preference profile with $P(M)$ for all men and $P(N)$ for all non-manipulators, there exists a profile for the manipulators $P(L)$ such that μ is the M-optimal stable matching for the total preference profile $(P(M), P(N), P(L))$, if and only if for every vertex v in the corre-*

sponding suitor graph $G(P(M), P(N), \mu)$, there exists a path from s to v (s is a virtual vertex in the graph). Moreover, if such $P(L)$ exists, it can be constructed in $O(n^2)$.

4.3. Pareto-optimal strategy profiles

We are now ready to combine the structures mentioned above to analyze permutation manipulations. Notice that eliminating more rotations results in weakly worse matchings for all men, which are also weakly better matchings for all women, since all men in the rotation are rejected and have to make proposals to the next woman in his preference list. Thus, a manipulator's objective is to eliminate as many rotations as possible by permuting their preference lists. Since there is no direct rotation elimination in the Gale-Shapley algorithm, we try to figure out what kind of rotations can be eliminated, i.e., after eliminating these rotations, the corresponding matching is in S_A.

We first analyze the structure of the set of all rotations. Rotations are not always exposed in a reduced table. Some rotations may become exposed only after other rotations are eliminated. Thus, we define the precedence relation between rotations and based on that, we introduce two concepts, *closed set* and *maximal rotations*.

Definition 4.8 (*Precedence*). A rotation $R_1 = (\mathcal{M}_1, \mathcal{W}_1, \mathcal{W}_1^r)$ is said to explicitly precede another $R_2 = (\mathcal{M}_2, \mathcal{W}_2, \mathcal{W}_2^r)$ if R_1 and R_2 share a common man m such that R_1 moves m from some woman to w and R_2 moves m from w to some other woman. Let the relation precede be the transitive closure of the explicit precedence relation, denoted by $\prec$. Also, $R_1 \sim R_2$ if neither $R_1 \prec R_2$ nor $R_2 \prec R_1$.

Definition 4.9 (*Closed set*). A set of rotations $\mathcal{R}$ is closed if for each $R \in \mathcal{R}$, any rotation R' with $R' \prec R$ is also in $\mathcal{R}$. A closed set $\mathcal{C}$ is minimal in a family of closed sets $\mathscr{C}$, if there is no other closed set in $\mathscr{C}$ that is a subset of $\mathcal{C}$. Moreover, define $CloSet(\mathcal{R})$ to be the minimal closed set that contains $\mathcal{R}$.

Definition 4.10 (*Maximal rotation*). Given a closed set of rotations $\mathcal{R}$, R is called a maximal rotation of $\mathcal{R}$ if no rotation $R' \in \mathcal{R}$ satisfying $R \prec R'$. Let $Max(\mathcal{R})$ denote the set of all the maximal rotations in $\mathcal{R}$. Furthermore, $\mathcal{R}$ is called a principle set if $Max(\mathcal{R})$ contains only one rotation.

Simply put, R_1 precedes R_2 if R_2 can only be exposed after R_1 is eliminated. A rotation R can only be exposed after all rotations preceding R are eliminated. Thus only closed sets can be validly eliminated. Also, a closed set of rotations $\mathcal{R}$ is uniquely determined by $Max(\mathcal{R})$. Therefore, given a closed set $\mathcal{R}$, the corresponding matching after eliminating rotations in $\mathcal{R}$ is determined by $Max(\mathcal{R})$.

The following theorem by Irving and Leather [1986] shows that closed sets of rotations are all that we need to consider.

THEOREM 4.11 (IRVING AND LEATHER [1986]). *Let S be the set of all stable matchings for a given preference profile, there is a one-to-one correspondence between S and the family of all closed sets.*

Then we focus on the changes made to the suitor graph when a rotation R is eliminated. We keep track of every proposal made by men in R and modify the graph accordingly. We first assume that the virtual vertex s is comparable with each man and for every $w \in W$ and every $m \in M$, $m \succ_w s$. When eliminating a rotation, we follow the steps below to modify the graph:

(1) Let all women in R reject their current partner, i.e., delete the edge (w_i, m_i) involved in R for each i;
(2) Arbitrarily choose a man m_i in R and let him propose to the next woman w in his preference list. Repeat until all man in R are accepted:

(a) If w is a manipulator, add an edge from m_i to w and delete edge (s, w) if it exists;
(b) If w is not a manipulator, then compare m_i with the two men (one is possibly s) in $V' = \{v \mid (v, w) \in E\}$. If m is not the worst choice, add an edge from m_i to w and delete the worst edge, and we say w is *overtaken* by m_i;
(c) If w accepts m_i, add an edge from w to m_i;

Let G and G' be the graphs corresponding the reduced tables before and after the elimination of R. It is easy to check that after modifying G using the steps defined above, the resulting graph is exactly G'. Before we explore the properties of the graph after eliminating rotations, we need to define a special structure in the suitor graph called strong components.

Definition 4.12. A sub-graph G' of G is said to be strongly connected if for every two vertices u and v in the G', there exists a path from u to v in G'. A strong component of a graph is a maximal strongly connected sub-graph.

The following lemma gives some connectivity properties of the suitor graph after eliminating a single rotation.

LEMMA 4.13. *After eliminating a rotation R,*

(*1*) *all agents in R are in the same strong component;*
(*2*) *vertices that are formerly reachable from a vertex in R remain reachable from R;*
(*3*) *vertices that are overtaken during the elimination of R are reachable from R.*

With Lemma 4.13, we do not need to worry about vertices that are reachable from vertices in R, for they will remain reachable after the elimination. Also, vertices that are overtaken and the other vertices reachable from overtaken vertices can be reached from vertices in R after the elimination.

In fact, every vertex is reachable from s in the initial suitor graph. Therefore, if a vertex becomes unreachable from s after eliminating a rotation, there must exist some edge that is deleted during the elimination, which only happens when some woman is overtaken. The next lemma extends Lemma 4.13 to a closed set of rotations.

LEMMA 4.14. *After eliminating a closed set of rotations $\mathcal{R}$, each v in $\mathcal{R}$ is reachable from at least one vertex in $Max(\mathcal{R})$, i.e., there exists a path to v from a vertex in $Max(\mathcal{R})$.*

This lemma is a generalization of Lemma 4.13. If R_2 explicitly precedes R_1, then they must contain a common man. Therefore, after eliminating R_1, vertices in R_1 can reach any vertex that is previously reachable from R_2. The analysis goes on recursively until some rotation has no predecessors.

Given a closed set of rotations $\mathcal{R}$, we say $\mathcal{R}$ can be eliminated for simplicity, if the corresponding matching after eliminating rotations in $\mathcal{R}$ is in S_A.

LEMMA 4.15. *A closed set of rotations $\mathcal{R}$ can be eliminated if and only if after eliminating $\mathcal{R}$, every vertex of rotations in $Max(\mathcal{R})$ can be reached from s.*

The following theorem provides us a desired property of closed sets of rotations that can be eliminated, based on which, we are able to design a polynomial time algorithm to find a Pareto-optimal matching.

THEOREM 4.16. *Given a closed set of rotations $\mathcal{R}$, if $\mathcal{R}$ can be eliminated, then there exists a rotation $R \in \mathcal{R}$ such that $CloSet(R)$ can be eliminated.*

Despite the fact that an optimal strategy profile may not exist, we know for sure that a Pareto-optimal strategy profile always exists since the set S_A is finite and nonempty. Algorithm 1 computes such a profile.

ALGORITHM 1: Find a Pareto-optimal profile for permutation manipulations

```
Find the set of all rotations $\mathcal{R}$ and all principle sets $\mathscr{P} = \{CloSet(R) | R \in \mathcal{R}\}$;
while True do
    Construct $\mathscr{C} = \{\mathcal{P} \in \mathscr{P} \mid \mathcal{P}$ can be eliminated$\}$;
    if $\mathscr{C} = \emptyset$ then
        Construct $P(L)$ for $L$ and return;
    else
        Arbitrarily choose a principle set $\mathcal{P}^* \in \mathscr{C}$ and eliminate $\mathcal{P}^*$;
```

In fact, for any iteration of Algorithm 1, the matching at the beginning of each iteration is in S_A. Therefore, if a closed set of rotations $\mathcal{R}$ can be eliminated, we can always find a principle set $\mathcal{P}^*$ contained in $\mathcal{R}$ such that $\mathcal{P}^*$ can be eliminated according to Theorem 4.16. To analyze the time complexity of Algorithm 1, we define a graph describing the precedence relation between rotations.

Definition 4.17 (*Precedence graph*). Given a set of rotations $\mathcal{R}$, let D be a directed acyclic graph, where the vertices in D are exactly $\mathcal{R}$, and there is an edge (R_1, R_2) in D if $R_1 \prec R_2$. Moreover, let H be the transitive reduction of D defined above, and H_r be the graph H with all edges reversed.

Note that H is exactly the directed version Hasse diagram of the precedence relation between rotations. For a rotation R, $CloSet(R)$ is the set of vertices that can be reached from R through a directed path in H_r. We split the algorithm into the initialization part and iteration part, and assume $|M| = |W| = n$.

In the initialization part, we first compute the initial matching using the Gale-Shapley algorithm, which can be computed in $O(n^2)$ time. Next we find all rotations with respect to preference profile P and also find all the principle sets. These two operations depend on the graph H_r. However, the graph H is the transitive reduction of D, and the construction of H is somewhat complex. Gusfield [1987] discusses how to find all rotations, whose number is $O(n^2)$, in $O(n^2)$ time. Instead of constructing H, Gusfield considered a sub-graph H' of D, whose transitive closure is identical to D. Moreover, H' can be constructed in $O(n^2)$ time. We will not discuss how to construct H' in detail but only apply Gusfield's results here. Then for each rotation R, we only need to search H' to find $CloSet(R)$, which takes $O(n^2)$ time. Thus, we finish the initialization step in $O(n^4)$ time since there are $O(n^2)$ rotations altogether.

The iteration part is the bottleneck of the algorithm. At least one rotation is eliminated for each iteration, and thus $O(n^2)$ iterations are needed. Inside each iteration, we need to construct the set $\mathscr{C}$. There are $O(n^2)$ principle sets and to determine whether a principle set can be eliminated, we need to simulate the Gale-Shapley algorithm and modify the suitor graph accordingly. After the modification, we traverse the suitor graph to see if each vertex is reachable. Both of the two operations takes $O(n^2)$ time. Thus, the construction of $\mathscr{C}$ takes $O(n^4)$ time. In the ***If-Else*** statement, if we find a principle set that can be eliminated, we eliminate the principle set and modify the suitor graph in $O(n^2)$. Otherwise, we traverse the suitor graph to construct the preference profile for L according to Theorem 4.7. Thus, the ***If-Else*** statement takes $O(n^2)$ time and totally, the time complexity is $O(n^6)$. To sum up, formally we have,

THEOREM 4.18. *Algorithm 1 correctly computes a Pareto-optimal strategy profile and the time complexity of Algorithm 1 is $O(n^6)$.*

Moreover, notice that at each iteration, the algorithm has multiple choices to select a principle set to eliminate. In fact, for each possible Pareto-optimal matching μ, there exists a way to select the principle sets such that the induced matching from the output of Algorithm 1 is μ.

THEOREM 4.19. *All Pareto-optimal strategy profiles can be found by the selection of principle sets inside each iteration of Algorithm 1.*

4.4. Inconspicuousness of feasible manipulations

In fact, in order to obtain a feasible matching, the manipulators only need slight modifications to their true preference lists. Formally,

THEOREM 4.20. *For each feasible manipulation, there exists a preference profile for the manipulators, in which each manipulator only needs to move at most one man in her true preference list to some higher ranking, that yields the same matching.*

For convenience, we introduce a new notation $Pro(w)$ for each $w \in W$. A proposal list $Pro(w)$ of a woman is a list of all men who have proposed to her in the Gale-Shapley algorithm, and the orderings of its entries are the same as her stated preference list. A reduced proposal list contains the top two entries (first entry if only one entry exists) of $Pro(w)$, denoted by $Pro_r(w)$. Clearly, each woman w is matched to the first man of $Pro_r(w)$. The following algorithm computes a Pareto-optimal strategy profile that is inconspicuous with respect to the true preference profile.

ALGORITHM 2: The inconspicuous version of Algorithm 1

```
Use Algorithm 1 to compute a strategy profile $P(L)$ for $L$;
Compute $Pro_r(w)$ for each $w \in L$ with respect to $P(L)$;
for each $w$ in $L$ do
    Modify the true preference list $P(w)$ by moving the second man in $Pro_r(w)$ to the position
    right after the first man in $Pro_r(w)$;
return the modified preference profile $P$;
```

4.5. Super-strong Nash equilibrium

The strategy profiles computed by Algorithm 2 not only are Pareto-optimal, but also form super-strong Nash equilibria.

THEOREM 4.21. *Under the feasibility assumption, the set of all Pareto-optimal matchings is the same as the set of super-strong Nash equilibrium outcomes.*

However, if we relax the feasibility assumption, the strategy profiles computed by Algorithm 2 do not necessarily form super-strong Nash equilibria. A counter-example is provided in Appendix C.2.

4.6. Hardness results

The above discussion shows that the Gale-Shapley algorithm is vulnerable to coalition manipulation. However, it is NP-complete to find a feasible manipulation in which every manipulator is strictly better off.

THEOREM 4.22 (THEOREM 1.3). *In permutation manipulations, it is NP-complete to find a manipulation such that the induced matching is strictly better off for all manipulators and stable with respect to true preference lists.*

Interestingly, our reduction can also be extended to search manipulations in unstable matchings due to the fact that, in the constructed stable matching problem, if all manipulators are better off in a matching, the matching must be stable.

THEOREM 4.23. *In permutation manipulations, it is NP-complete to find a manipulation such that the induced matching (not necessarily stable with respect to the true preference profile) is strictly better off for all manipulators.*

According to Theorem 4.22, one immediate corollary is that the number of different Pareto-optimal matchings cannot be polynomial in the number of men and women. For otherwise, we can enumerate all such matchings by Algorithm 1 to develop a polynomial time algorithm. Actually, there are $2^{\Theta(|M|+|W|)}$ different induced matchings, which is Pareto-optimal and weakly better off for all manipulators, in the constructed stable matching problem used in the reduction of Theorem 4.22 (see Appendix D for details). Last but not least, to compute the number of matchings that are Pareto-optimal and strictly better off for all manipulators is #P-Hard.

THEOREM 4.24. *In permutation manipulations, it is #P-Hard to compute the number of induced matchings, which are strictly better off, Pareto-optimal for all manipulators, and stable with respect to true preference lists.*

5. CONCLUSION AND FUTURE WORKS

In this paper, we consider two types of manipulations, general manipulations and permutation manipulations, by any subset of women in the Gale-Shapley algorithm.

- In general manipulations, we show that the induced manipulation game always has a strong Nash equilibrium, which is also unique and Pareto-dominant.
- In permutation manipulations, we present a polynomial-time algorithm to find a Pareto-optimal strategy profile and prove that Pareto-optimal matchings are equivalent to super-strong Nash equilibrium outcomes.

Along with theoretical results, we introduce some new concepts, such as maximum rotation and principle set, and develop novel techniques to connect rotations and suitor graphs, which are useful for further study. Moreover, although these results show the vulnerabilities of the Gale-Shapley algorithm, we also show that finding a manipulation that induces a matching strictly better off for all manipulators is NP-complete.

Most of the results in this paper are obtained under the feasibility assumption, especially in the permutation manipulation setting. Finding Nash equilibria or Pareto-optimal matchings without the feasibility assumption would still be interesting. It remains unknown whether finding such solutions in this setting is easy or not.

REFERENCES

Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. 2005. The new york city high school match. *American Economic Review* (2005), 364–367.

Atila Abdulkadiroglu and Tayfun Sönmez. 2003. School choice: A mechanism design approach. *The American Economic Review* 93, 3 (2003), 729–747.

David J Abraham, Avrim Blum, and Tuomas Sandholm. 2007. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, 295–304.

Itai Ashlagi, Yashodhan Kanoria, and Jacob Leshno. 2013. Unbalanced Random Matching Markets. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce (EC '13)*. ACM, New York, NY, USA, 27–28. DOI:http://dx.doi.org/10.1145/2482540.2482590

Haris Aziz, Hans Georg Seedig, and Jana Karina von Wedel. 2015. On the Susceptibility of the Deferred Acceptance Algorithm. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 939–947.

Gabrielle Demange, David Gale, and Marilda Sotomayor. 1987. A further note on the stable matching problem. *Discrete Applied Mathematics* 16, 3 (1987), 217–222.

Lester E Dubins and David A Freedman. 1981. Machiavelli and the Gale-Shapley algorithm. *American mathematical monthly* (1981), 485–494.

D. Gale and L. S. Shapley. 1962. College Admissions and the Stability of Marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15.

David Gale and Marilda Sotomayor. 1985. Ms. Machiavelli and the stable matching problem. *Amer. Math. Monthly* (1985), 261–268.

Allan Gibbard. 1973. Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society* (1973), 587–601.

Yannai A. Gonczarowski. 2014. Manipulation of Stable Matchings Using Minimal Blacklists. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation (EC '14)*. ACM, New York, NY, USA, 449–449. DOI:http://dx.doi.org/10.1145/2600057.2602840

Yannai A Gonczarowski and Ehud Friedgut. 2013. Sisterhood in the Gale-Shapley Matching Algorithm. *The Electronic Journal of Combinatorics* 20, 2 (2013), P12.

Dan Gusfield. 1987. Three fast algorithms for four problems in stable marriage. *SIAM J. Comput.* 16, 1 (1987), 111–128.

Dan Gusfield. 1988. The structure of the stable roommate problem: efficient representation and enumeration of all stable assignments. *SIAM J. Comput.* 17, 4 (1988), 742–769.

Dan Gusfield and Robert W. Irving. 1989. *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, Cambridge, MA, USA.

Chien-Chung Huang. 2006. Cheating by men in the Gale-Shapley stable matching algorithm. In *Algorithms–ESA 2006*. Springer, 418–431.

Chien-Chung Huang. 2010. Classified stable matching. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 1235–1253.

Nicole Immorlica and Mohammad Mahdian. 2005. Marriage, honesty, and stability. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 53–62.

Robert W Irving. 1985. An efficient algorithm for the stable roommates problem. *Journal of Algorithms* 6, 4 (1985), 577–595.

Robert W Irving and Paul Leather. 1986. The complexity of counting stable marriages. *SIAM J. Comput.* 15, 3 (1986), 655–667.

Robert W Irving and David F Manlove. 2009. Finding large stable matchings. *Journal of Experimental Algorithmics (JEA)* 14 (2009), 2.

Robert W Irving, David F Manlove, and Sandy Scott. 2000. The hospitals/residents problem with ties. In *Algorithm Theory-SWAT 2000*. Springer, 259–271.

Paula Jaramillo, Çaatay Kayı, and Flip Klijn. 2014. On the exhaustiveness of truncation and dropping strategies in many-to-many matching markets. *Social Choice and Welfare* 42, 4 (2014), 793–811.

Donald E. Knuth. 1976. *Mariages Stables*. Les Presses de I'Universite de Montreal.

Donald E Knuth, Rajeev Motwani, and Boris Pittel. 1990. Stable husbands. In *Proceedings of the first annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 397–404.

Hirotatsu Kobayashi and Tomomi Matsui. 2009. Successful manipulation in stable marriage model with complete preference lists. *IEICE TRANSACTIONS on Information and Systems* 92, 2 (2009), 116–119.

Hirotatsu Kobayashi and Tomomi Matsui. 2010. Cheating strategies for the Gale-Shapley algorithm with complete preference lists. *Algorithmica* 58, 1 (2010), 151–169.

Jinpeng Ma. 2010. The singleton core in the college admissions problem and its application to the National Resident Matching Program (NRMP). *Games and Economic Behavior* 69, 1 (2010), 150–164.

David G McVitie and Leslie B Wilson. 1971. The stable marriage problem. *Commun. ACM* 14, 7 (1971), 486–490.

Maria Silvia Pini, Francesca Rossi, K Brent Venable, and Toby Walsh. 2009. Manipulation and gender neutrality in stable marriage procedures. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 665–672.

Alvin Roth and Marilda Oliveira Sotomayor. 1992. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press.

AlvinE. Roth and JohnH. Vande Vate. 1991. Incentives in two-sided matching with random stable mechanisms. *Economic Theory* 1, 1 (1991), 31–44. DOI:http://dx.doi.org/10.1007/BF01210572

Alvin E Roth. 1982. The economics of matching: Stability and incentives. *Mathematics of operations research* 7, 4 (1982), 617–628.

Alvin E Roth. 1986. On the allocation of residents to rural hospitals: a general property of two-sided matching markets. *Econometrica: Journal of the Econometric Society* (1986), 425–427.

Alvin E Roth. 1996. The national residency matching program as a labor market. *Journal of the American Medical Association* 275, 13 (1996), 1054–1056.

Alvin E Roth. 2002. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica* 70, 4 (2002), 1341–1378.

Alvin E Roth, Tayfun Sönmez, and M Ünver. 2004. Kidney exchange. *The Quarterly Journal of Economics* 119, 2 (2004), 457–488.

Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. 2005. A kidney exchange clearinghouse in New England. *American Economic Review* (2005), 376–380.

Noriyoshi Sukegawa and Yoshitsugu Yamamoto. 2012. Preference profiles determining the proposals in the Gale–Shapley algorithm for stable matching problems. *Japan journal of industrial and applied mathematics* 29, 3 (2012), 547–560.

Chung-Piaw Teo, Jay Sethuraman, and Wee-Peng Tan. 2001. Gale-shapley stable marriage problem revisited: Strategic issues and applications. *Management Science* 47, 9 (2001), 1252–1267.

APPENDIX

A. OMITTED PROOFS IN SECTION 3

A.1. Proof of Lemma 3.1

PROOF. Let μ be any matching in S_A and $P(L)$ be the corresponding reported preference profile by the manipulators in a Nash equilibrium of general manipulation. Therefore, μ is the M-optimal matching of $P' = (P(M), P(N), P(L))$. We construct a truncated preference profile $P_t(L)$ for the manipulators where for each manipulator w, $\mu(w)$ is the last in her preference lists. (If w is single in μ, her preference list remains the same as her true preference list).

Notice that μ is stable under the true preference profile. We show that μ is also stable under $P_t = (P(M), P(N), P_t(L))$. Clearly, μ is individually rational. Assume on the contrary that a pair (m, w) blocks μ. Then we must have $m \succ_w^{P_t} \mu(w)$ and $w \succ_m^{P_t} \mu(m)$. From the construction of P_t, we know that $m \succ_w^{P} \mu(w)$ is also true $\forall w \in W$. Also, $w \succ_m^{P} \mu(m)$ is true since no man's preference list is changed. Thus, (m, w) is a blocking pair under P, which contradicts to the stability of μ.

We claim that μ is the W-pessimal matching under P_t. Otherwise, assume μ^* is the W-pessimal matching and $\mu \neq \mu^*$. For each manipulator w, we have $\mu(w) = \mu^*(w)$, for $\mu(w)$ is already the last man in her preference list. We consider μ^* under P'. Since all manipulators are matched to the same men in the two matchings and $\mu \neq \mu^*$, there must be a non-manipulator w' such that $\mu(w') \succ_{w'}^{P'} \mu^*(w')$. Thus, μ^* is not stable under P' since μ is the W-pessimal matching under P' and there is a blocking pair (m'', w''). Moreover, w'' cannot be a manipulator, otherwise since $\mu(m'') = \mu^*(m'')$ (implied by $\mu(w'') = \mu^*(w'')$) and $m'' \succ_{w''}^{P'} \mu(m'')$, m'' is in $P_t(w'')$ so that (m'', w'') blocks μ^* under P_t. As a result, w'' is a non-manipulator but (m'', w'') also blocks μ^* under P_t, since the preference lists of both m and w are the same in the two preference profiles. A contradiction. □

A.2. Proof of Theorem 3.4

In order to prove Theorem 3.4, we first consider the following lemma.

LEMMA A.1. *Let $(P(M), P(W))$ be the true preference profile and $P(L)$ be the reported profile by the manipulators. If for each manipulator w, her W-optimal partner is not removed from her list in truncation manipulation, then $S(P(M), P(N), P(L)) \subseteq S(P(M), P(W))$.*

PROOF. Suppose not and there exists a matching μ which is in $S(P(M), P(N), P(L))$ but not in $S(P(M), P(W))$. Then, there exists a blocking pair (m, w) in μ under true preference lists. For m, since his preference list is not modified, he prefers w to $\mu(m)$ in true preference lists and the lists after truncation.

If w is not single in μ, then m is still in her preference list after truncation manipulation since $m \succ_w \mu(w)$, which forms a blocking pair in μ. Otherwise, if w is single in μ, notice that, since the order of each man and each woman's preference list is not changed, W-optimal matching is in $S(P(M), P(N), P(L))$. Thus, w is single in W-optimal matching and according to the assumption, she is not a manipulator. □

PROOF OF THEOREM 3.4. We first prove that each manipulator is matched to her M-optimal partner if they report $P(L)$. According to Lemma A.1 the induced matching μ must be in $S(P(M), P(W))$. Notice that, W-optimal matching is still in $S(P(M), P(N), P(L))$. Thus, according to Theorem 3.3, each manipulator is not single after manipulation, and she cannot be matched with a man worse than her W-optimal partners since she already removes him. Also, each woman cannot get a partner better than her W-optimal partner. Thus, all manipulators must be matched with their W-optimal partner.

Next we show that it is a super-strong Nash equilibrium for all manipulators to do so. Suppose that there exists a sub-coalition of women $L' \subseteq L$ who can deviate from $P(L)$ so that each manipulator is weakly better off and at least one is strictly better off. However, the induced matching must be stable under true preference lists, but all manipulators are matched with their W-optimal partner already. A contradiction. $\square$

A.3. Proof of Theorem 3.7

Before we discuss the lattice structure of the set of S_A, we first prove a lemma about the relation between the length of the preference lists and the induced matching. We say a preference profile $P_1 = (P(M), P(N), P_1(L))$ is shorter than another $P_2 = (P(M), P(N), P_2(L))$ if for each $w \in L$, $|P_1(w)| \leq |P_2(w)|$. In other words, P_1 is shorter than P_2 if all manipulators remove no less men in P_1 than in P_2.

LEMMA A.2. *Let P_1 and P_2 be two preference profiles, and μ_1 and μ_2 be the two corresponding M-optimal matchings, if for each manipulator, her W-optimal partner is in both P_1 and P_2, and P_1 is shorter than P_2, then $\mu_1 \succeq_W \mu_2$.*

PROOF. The lemma is a corollary of Lemma A.1. Since P_1 is shorter than P_2, P_1 can be viewed as a manipulation starting from P_2. By Lemma A.1, the set of stable matchings under P_1 is a subset of that under P_2. Moreover, the Gale-Shapley outputs the W-pessimal matching and thus, $\mu_1 \succeq_W \mu_2$. $\square$

LEMMA A.3. *Given two preference profiles P_1 and P_2, and two corresponding M-optimal matchings by μ_1 and μ_2. Let $P_\cap = P_1 \cap P_2 = (P(M), P(N), P_\cap(L))$ such that*

$$P_\cap(w) = \begin{cases} P_1(w) & \textit{if } |P_1(w)| \leq |P_2(w)| \\ P_2(w) & \textit{otherwise} \end{cases}$$

for all $w \in L$. Then the M-optimal matching $\mu_\cap$ under $P_\cap$ is exactly $\mu_\wedge = \mu_1 \wedge \mu_2$.

PROOF. It is easy to check that $P_\cap(L)$ is a legal profile for the manipulators, i.e., the preference list for each manipulator w can be obtained by truncating her true preference list.

According to Lemma A.2, we have $\mu_\cap \succeq_W \mu_1$ and $\mu_\cap \succeq_W \mu_2$, so that from the definition of $\mu_\wedge$, we have $\mu_\cap \succeq_W \mu_\wedge$. To show that $\mu_\wedge$ is identical to $\mu_\cap$, we only need to show that $\mu_\wedge \succeq_W \mu_\cap$. We claim that $\mu_\wedge$ is a stable matching under $P_\cap$, and thus $\mu_\wedge \succeq_W \mu_\cap$ because $\mu_\cap$ is the W-pessimal matching under $P_\cap$.

For each w, $\mu_\wedge(w) \succeq_w \mu_1(w)$, so $\mu_\wedge(w)$ is in $P_1(w)$. Similarly $\mu_\wedge(w)$ is also in $P_2(w)$. Therefore $\mu_\wedge(w)$ is individually rational under $P_\cap$. Assume that $\mu_\wedge$ is not stable under $P_\cap$. Then there must be a blocking pair (m, w) and $m \succ_w \mu_\wedge(w)$ and $w \succ_m \mu_\wedge(m)$ under $P_\cap$. However, the two inequalities also hold in both $P_1(w)$ and $P_2(w)$. Notice that (m, w) is unmatched in at least one of the two matchings μ_1 and μ_2, otherwise $\mu_\wedge(w) = m$. Thus, (m, w) blocks either μ_1 or μ_2, which produces a contradiction. $\square$

However, it is not a meet-semilattice. Consider the counter-example shown in Table II. The matching results are in Table III.

Suppose $L = \{w_1, w_2\}$. If w_1 alone lies and cuts her list to the one containing only m_5, the induced matching is μ_1. If w_2 alone lies and lists only m_6, we will get μ_2. But the meet of the these two matchings $\mu_\vee$ cannot be induced by only truncating the preference lists of w_1 and w_2.

Nevertheless, we prove that the set of partial matchings S_A^L is a lattice.

Table II. A counter example for the meet-semilattice structure

Men's preference lists						
m_1	w_1	w_3	w_5	$-$	$-$	$-$
m_2	w_2	w_3	w_6	$-$	$-$	$-$
m_3	w_3	w_4	$-$	$-$	$-$	$-$
m_4	w_4	w_3	$-$	$-$	$-$	$-$
m_5	w_5	w_1	$-$	$-$	$-$	$-$
m_6	w_6	w_2	$-$	$-$	$-$	$-$

Women's preference lists						
w_1	m_5	m_1	$-$	$-$	$-$	$-$
w_2	m_6	m_2	$-$	$-$	$-$	$-$
w_3	m_4	m_2	m_1	m_3	$-$	$-$
w_4	m_3	m_4	$-$	$-$	$-$	$-$
w_5	m_1	m_5	$-$	$-$	$-$	$-$
w_6	m_2	m_6	$-$	$-$	$-$	$-$

Table III. Induced matchings and their meet

μ_1		μ_2		$\mu_\vee$	
m_1	w_5	m_1	w_1	m_1	w_1
m_2	w_2	m_2	w_6	m_2	w_2
m_3	w_4	m_3	w_4	m_3	w_4
m_4	w_3	m_4	w_3	m_4	w_3
m_5	w_1	m_5	w_5	m_5	w_5
m_6	w_6	m_6	w_2	m_6	w_6

LEMMA A.4. *Given $P_1 = (P(M), P(N), P_1(L))$ and $P_2 = (P(M), P(N), P_2(L))$, suppose μ_1 and μ_2 are the two corresponding M-optimal matchings. Let $\mu_\vee = \mu_1 \vee \mu_2$. Then $\mu_\vee^L$ is in S_A^L.*

PROOF. We construct a preference profile $P_\cup$ as follows. For each $w \in L$, she removes all men ranked below $\mu_\vee(w)$ in her true preference list. We prove that the corresponding M-optimal matching $\mu_\cup$ is an extension of $\mu_\vee^L$, i.e., $\mu_\cup^L = \mu^L$.

Using similar techniques in the proof of Lemma A.3, we conclude that $\mu_\vee$ is a stable matching with respect to preference profile $P_\cup$. For each $w \in L$, since $\mu_\vee(w)$ is the last one in her preference list and they must be matched to their W-pessimal partner under $P_\cup$, $\mu_\vee(w)$ must be equal to $\mu_\cup(w)$ and $\mu_\vee^L = \mu_\cup^L$. □

Lemma A.3 is clearly true when restricted to manipulators. Combining the above two lemmas together, we immediately get Theorem 3.7.

A.4. Proof of Theorem 3.8

This theorem is a direct corollary of the following theorem.

THEOREM A.5 (LIMITS ON SUCCESSFUL MANIPULATION, DEMANGE ET AL. [1987]). *Let P be the true preferences (not necessarily strict) of the agents, and let P' differ from P in that some coalition C of men and women mis-state their preferences. Then there is no matching μ, stable for P', which is strictly preferred to every stable matching under the true preferences P by all members of C.*

PROOF OF THEOREM 3.8. In proof of Theorem 3.4, we have already shown that for all stable matchings in $S(P(M), P(N), P(L))$, each manipulators are matched with their W-optimal partner. Applying Theorem A.5 with $C \subseteq L$, we can conclude that there is no matching in $S(P(M), P(N), P(L\backslash C), P(C))$ is strictly preferred to every stable matching in $S(P(M), P(N), P(L))$ for all members of C. Therefore, the constructed strategy profile is a strong Nash equilibrium. □

B. OMITTED PROOFS IN SECTION 4

B.1. Proof of Lemma 4.5

PROOF. We use Irving's algorithm to solve the corresponding SR problem. It is straightforward to check that at the end of phase 1, each man is engaged to his M-optimal partner of the origin SM problem and each woman is engaged to her W-optimal partner. It follows that the reduced list of each agent is composed of only agents of the opposite sex. Therefore, any rotation must contain only agents of the same sex, since the reduced lists of two adjacent agents in a rotation share a common agent.

When we eliminate a rotation only containing men, each man of the rotation will be engaged to the second woman in his reduced list and thus become worse off. Assume that not all rotations have dual rotations, i.e., there exists a singleton rotation R. Without loss of generality, suppose R is a rotation only containing men and m is a man contained in R. Then for any solution of the problem, R must be eliminated to generate that solution. Thus, m cannot be matched to his M-optimal woman at the end of phase 2, which contradicts to the fact that the M-optimal matching is a solution. □

B.2. Proof of Lemma 4.13

The proof is based on the following lemmas.

LEMMA B.1. *For each man m_i in R, in the procedure of eliminating the rotation R, w_{i+1} (the subscript is taken modulo r) is the first woman to accept him, and each woman in R accepts only one proposal during the procedure.*

PROOF. According to the definition of rotations, w_{i+1} is the second in m_i's reduced list. If there are other women between w_i and w_{i+1} in m_i's preference list, they are absent from the reduced list because these women already hold proposals from better men. Henceforth, even though m_i proposes to these women, they reject him. But m_i is in w_{i+1}'s reduced list since w_{i+1} is in m_i's. Therefore, m_i is a better choice for w_{i+1} and w_{i+1} accepts him.

After the elimination, each man m_i in R proposes to w_{i+1} and each man is accepted only once. Also each woman w_{i+1} holds a new proposal from m_i and thus accepts at least once. The conclusion is immediate since the total number of each man being accepted equals to the total number of each woman accepting a new partner. □

LEMMA B.2. *After eliminating a rotation $R = (\mathcal{M}, \mathcal{W}, \mathcal{W}^r)$, all agents in R are in the same strong component.*

PROOF. For each m_i in R, R moves m_i from w_i to w_{i+1}. As a result, there exists an edge from w_{i+1} to m_i. We now prove that each m_i has an outgoing edge pointing to w_i, and all agents in R then form a cycle, and thus in the same strong component. Before the elimination, w_i is the partner of m_i, so there is an edge from m_i to w_i. If w_i is a manipulator, the edge (m_i, w_i) is not removed during the elimination according to the steps described above. If w_i is not a manipulator, then only two incoming edges are remained after the elimination and these edges are from the best two men among those who propose to her. According to Lemma B.1, only one man, namely m_{i-1}, is accepted. Thus, m_{i-1} is the best suitor of w_i. We claim that m_i is the second best and the edge from m_i is still in the suitor graph. Otherwise, suppose m' is a better choice than m_i to w_i. Then m' is also in R. We let m' propose first, and w_i accepts m', which makes w_i accepts at least twice. A contradiction. □

From the proof of Lemma B.2 we know that if v is in R and is a non-manipulator, then after eliminating R, the two incoming edges are both from inside the rotation. Only manipulators may have edges coming from outside of the rotation.

PROOF OF LEMMA 4.13. Since each woman can be reached from her partner before the elimination, it is without loss of generality to assume that a vertex v can be reached from a man m in R through a path p. Let u be the last vertex in p such that u is in R or is overtaken by a vertex in R. If u is in R, then after the elimination, m can reach u since they are in the same strong component. If u is overtaken by some vertex m', then during the elimination, an edge (m', u) is added to the graph. Thus, m can reach u through m'. Henceforth, in any case, u is reachable. Since in p the vertices between u

and v are neither in R nor overtaken by some vertex in R, the path from u to v remains in the modified graph. Therefore v is reachable from m and also from any vertex in R for they are in the same strong component after the elimination. □

B.3. Proof of Lemma 4.14

PROOF. We eliminate the rotations in $\mathcal{R}$ one by one and generate a sequence of rotations $q = (R_1, R_2, \ldots, R_n)$. R_i is the i-th rotation to eliminate, and after eliminating R_n, all rotations in $\mathcal{R}$ are eliminated. Denote q_i as the set of the rotations before R_i in q. For each i, q_i is a closed set. We call i the sequence number of q_i and we prove by induction on the sequence number that after eliminating q_i, all vertices in q_i can be reached from a vertex in $Max(q_i)$. For $i = 1$, $q_i = \{R_1\}$, the case is trivial from Lemma B.2. Assume the statement is true for $i = k$, then for $i = k + 1$, we only eliminate one more rotation R_{k+1} than in the case with $i = k$. R_{k+1} is in $Max(q_{k+1})$ otherwise there exists another rotation R' in q_k such that $R_{k+1} \prec R'$ and then q_k is not a closed set. Let $D = Max(q_k) \setminus Max(q_{k+1})$. Rotations in D are no longer maximal rotations simply because R_{k+1} is eliminated, which indicates that rotations in D explicitly precede R_{k+1}. Henceforth, every rotation R in D has a common agent with R_{k+1} and each vertex u reachable from R is reachable from that common agent. According to Lemma 4.13, u can be reached from R_{k+1}. For each vertex u' that is not reachable from rotations in D, it must be reachable from another rotation R' in $Max(q_k)$ through path p and R' is still in $Max(q_{k+1})$. If p is still in the suitor graph, then we are done. Otherwise, some vertices in p must be in R_{k+1} or is overtaken by a man in R_{k+1}. Let z be the last vertex in p such that z is in R_{k+1} or is overtaken. z can be reached from R_{k+1} and the path from z to u' is not affected by the elimination. Therefore, u' is reachable from R_{k+1}. □

B.4. Proof of Lemma 4.15

PROOF. If a closed set of rotations $\mathcal{R}$ can be eliminated, then every vertex is reachable after $\mathcal{R}$ is eliminated. As a result, any member of $Max(\mathcal{R})$ is reachable.

If after eliminating $\mathcal{R}$, any member of $Max(\mathcal{R})$ can be reached from s, then we need to show that all other vertices are also reachable from s. We split all vertices into two parts. Let V denote the set of all the vertices that can be reached from members of $Max(\mathcal{R})$. If a vertex v is in V, then v is reachable from s through $Max(\mathcal{R})$. If v is not in V, then in the initial graph, there is a path p from s to v. We claim that all the vertices in path p is not in any of the rotations in $\mathcal{R}$ or overtaken when eliminating a rotation. Otherwise, according to Lemma 4.14, v is reachable from $Max(\mathcal{R})$. Thus, the path p is still in the graph after eliminating all the rotations in $\mathcal{R}$. □

B.5. Proof of Theorem 4.16

We first consider the following lemma about the maximal rotations of a closed set that can be eliminated.

LEMMA B.3. *If a closed set $\mathcal{R}$ can be eliminated, then every rotation in $Max(\mathcal{R})$ must contain a manipulator.*

PROOF. Assume on the contrary that there exists a rotation R in $Max(\mathcal{R})$ such that R contains no manipulators. We can always change the order of elimination to make R the last to eliminate. We prove that after eliminating R, any vertex in R is not reachable from s. From the proof of Lemma B.2, we know that all vertices in R form a cycle after eliminating R. Each man in R has only one incoming edge from his current partner who is also in R. Each woman has two incoming edges, one from her partner in R and another from her former partner which is also in R. Thus, every vertex in R has no incoming edges from outside the cycle and thus is not reachable from s. □

PROOF OF THEOREM 4.16. Let V be the set of all vertices in $\mathcal{R}$. After eliminating $\mathcal{R}$, we arbitrarily choose a vertex v in V. In the corresponding suitor graph, there is a path $p = (v_0 = s, v_1, v_2, \ldots, v_n = v)$ from s to v since $\mathcal{R}$ can be eliminated. Let u be the first vertex in p such that u is in V. u is obviously not v_1, or otherwise the edge (s, u) will be deleted. Moreover, u must be in L, since any non-manipulator can only be reached from a node in V if she is overtaken during the elimination. Assume $u = v_l$ and $l > 1$. Then the sub-path $p' = (v_0, v_1, \ldots, v_l = u)$ is not affected (no vertices in V or overtaken) during the elimination. Henceforth, p' is in the original suitor graph before eliminating $\mathcal{R}$. Now we consider the set $\mathcal{R}' = \{R \in \mathcal{R} | u \in R\}$. For any R in $\mathcal{R}'$, if we eliminate $CloSet(R)$, the sub-path is also not affected. Therefore $CloSet(R)$ can be eliminated according to Lemma 4.15. □

B.6. Proof of Theorem 4.18

The remaining thing to prove is the correctness of the algorithm. We begin with the following lemma.

LEMMA B.4. *Given a set of manipulators $L \in W$, and the true preference profile $P = (P(M), P(W))$. Let μ be any matching in S_A and $\mathcal{R}$ be the corresponding closed set of rotations. Then there exists a preference profile $P_\mu(L)$ for L such that μ is the M-optimal stable matching of the preference profile $P_\mu = (P(M), P(N), P_\mu(L))$, and the reduced table of P after eliminating $\mathcal{R}$ is exactly the reduced table of P_μ before eliminating any rotation.*

PROOF. Since μ is in S_A, there exists $P' = (P(M), P(N), P'(L))$ such that the induced matching μ. For each $w \in L$, we modify $P'(w)$ as follows:

(1) delete all men m such that $m \succ_w^P \mu(w)$;
(2) reinsert them at the beginning according to their order in w's true preference list;
(3) move $\mu(w)$ to the position right after all men m such that $m \succ_w^P \mu(w)$;

Denote the modified preference profile by P'_μ. In fact, P'_μ is the P_μ we are looking for.

We first prove that μ is the M-optimal matching under P'_μ. After the first two steps of modifications, the M-optimal matching is still μ, since for each w, we only change the position of men ranked higher than $\mu(w)$ in her true preference list, who must have not proposed to w under P', and thus do not change the output of the Gale-Shapley algorithm. Otherwise, if a man m with $m \succ_w^P \mu(w)$ has proposed to w, then we must have $w \succ_m^{P'} \mu(m)$, which is equivalent to $w \succ_m^P \mu(m)$. Thus (m, w) forms a blocking pair in μ under the true preference profile P, contradicting to the stability of μ under P. In the third step, we move $\mu(w)$ to the position right after all men ranked higher than $\mu(w)$ in the true preference list $P(w)$. Consider all the men m' with $m' \succ_w^{P'} \mu(w)$ but $\mu(w) \succ_w^{P'_\mu} m'$. m' must have not proposed to w under P', or otherwise $\mu(w)$ cannot be the partner of w. Therefore, the positions of the men in P'_μ do not affect the output of the Gale-Shapley algorithm.

Let T_{P_μ} be the reduced table of P after eliminating $\mathcal{R}$ and $T_{P'_\mu}$ be the reduced tables of P'_μ. We already know that for each woman, her partners in the two reduced tables are the same, which is $\mu(w)$. In fact, a change of reduced table happens if and only if a woman accepts a proposal from a man m and removes everyone ranked below m in her preference list. Henceforth, in the reduced list of each woman, no man is ranked below her current partner. Therefore, to prove that T_{P_μ} is the same as $T_{P'_\mu}$, if suffices to show that for each woman, P and P'_μ are the same after removing all men ranked below her current partner, which is clear from the construction of P'_μ. □

From Theorem 4.11, we know that only closed sets need to be considered. Although the Lemma B.3 has already ruled out all closed sets that have a maximal rotation containing only non-manipulators, there are still exponentially many possibilities. However, Theorem 4.16 shows that every closed set that can be eliminated contains a principle set, which can also be eliminated. A natural idea is to iteratively grow the closed set by adding principle sets. The above lemma shows that after each iteration, we can construct a problem that has the current matching as its initial matching, and contains rotations that are not yet eliminated. If we find a principle set that can be eliminated in the constructed problem, it can also be eliminated in the original problem.

PROOF OF THEOREM 4.18. Algorithm 1 can be summarized as iteratively eliminating principle sets. Inside each iteration, we scan all rotations to find closed sets that can be eliminated. The algorithm terminates since there are finite rotations. Let $P = (P(M), P(W))$ be the original preference profile and S be the set of all stable matchings in terms of P. Let $\mathcal{R}$ be the set of all rotations.

Assume on the contrary that the algorithm terminates with a strategy profile P_μ that is not Pareto-optimal. Denote μ as the matching produced by P_μ and let $\mathcal{R}_\mu$ be the corresponding set of rotations. Then there must be another strategy profile P^* that dominates P_μ. Suppose μ^* is the induced matching of P^* and $\mathcal{R}^*$ is the corresponding closed set of rotations.

According to Lemma B.4, we can construct a problem with the original preference profile P' such that μ is the M-optimal matching of P' and the reduced table of P after eliminating $\mathcal{R}_\mu$ is exactly the reduced table of P' before eliminating any rotation. It follows that the set of all rotations of P' is exactly $\mathcal{R}' = \mathcal{R} \setminus \mathcal{R}_\mu$ and the set of all stable matchings of P' is $S' = \{\mu' \in S(P(M), P(W)) | \mu' \succeq_W \mu\}$. Henceforth, μ^* is also in $S_A(P')$. Since P^* dominates P_μ, we have $\mathcal{R}_\mu \subset \mathcal{R}^*$ and $\mathcal{R}^* \setminus \mathcal{R}_\mu \subseteq \mathcal{R}'$. Therefore, according to Theorem 4.16, there exists a closed set $CloSet(R)$ that can be eliminated in P'. Let μ' be the induced matching after eliminating $CloSet(R)$ in P' and $P_{\mu'}(L)$ be a preference profile for the manipulators such that the induced matching is μ', i.e., μ' is the M-optimal matching of profile $P_{\mu'} = (P(M), P(N), P_{\mu'}(L))$. As a result, μ' is in $S_A(P)$ with corresponding profile $P_{\mu'}(L)$. Thus, after eliminating $\mathcal{R}_\mu$, $CloSet(R)$ can still be eliminated, which contradicts to the termination of the algorithm. □

B.7. Proof of Theorem 4.19

PROOF. Assume the $P(L)$ is a Pareto-optimal strategy profile for the manipulators. Let μ be the matching produced by $P(L)$ and $\mathcal{R}_\mu$ the corresponding set of rotations. μ can be forced to be the induced matching by always choosing the principle set that is a subset of $\mathcal{R}_\mu$. Let $\mathcal{R}_k$ be the rotations eliminated so far at the end of the k-th iteration and μ_k be the corresponding matching . We prove by induction on the iterations that at the end of each iteration, $\mathcal{R}_k$ is a subset of $\mathcal{R}_\mu$. In the first iteration, $\mathcal{R}_\mu$ is in S_A, so there exists a principle set $\mathcal{P} \subset \mathcal{R}_\mu$ that can be eliminated. Assume the statement holds for the k-th iteration. At the beginning of the $(k+1)$-th iteration, μ_k is the induced matching, and $\mathcal{R}_k$ is a subset of $\mathcal{R}_\mu$ by the inductive hypothesis, then there exists at least one principle set $\mathcal{P}_{k+1} \subset \mathcal{R}_\mu \setminus \mathcal{R}_k$ that can be eliminated. Therefore, at the end of the $(k+1)$-th iteration, $\mathcal{R}_{k+1} = \mathcal{R}_k \cup \mathcal{P}_{k+1}$ is also a subset of $\mathcal{R}_\mu$. When the algorithm terminates, the set of all eliminated rotations $\mathcal{R}$ is also a subset of $\mathcal{R}_\mu$. Assume $\mathcal{R} \neq \mathcal{R}_\mu$, then we can find some principle set to eliminate, which contradicts to the termination condition of the algorithm. Therefore the Pareto-optimal strategy profile can be found by the algorithm. □

B.8. Proof of Theorem 4.20

LEMMA B.5. *Given all agents' true preference profile $(P(M), P(W))$, if a matching μ is in S_A with corresponding preference profile $P = (P(M), P(N), P(L))$, then the induced matching is still μ, if for each $w \in L$, we modify w's preference list by moving $Pro_r(w)$ to the top and ordering other men arbitrarily.*

PROOF. Suppose the corresponding matching to the modified preference profile is μ'. We show that $\mu' = \mu$.

Let P and P' be the original profile and the modified profile. All the partial orders we used in this proof is defined in P. We construct a graph T, which is a sub-graph of suitor graph $G(P(M), P(N), \mu)$, according to the set of all reduced proposal lists in P. The set of vertices is just $M \cup W$, and the edges are $E = \{(w, \mu(w)) \mid w \in W\} \cup \{(m, w) \mid w \succ_m \mu(m), m \in Pro_r(w)\}$. We also add a virtual vertex s, and add edges from s to each woman who has no incoming edges. Note that every woman has an outgoing edge pointing to her mate in μ, and at most one incoming edge from her second entry in her proposal list. It is easy to prove that at least one woman has only one entry in her proposal list, and thus this woman has no incoming edge except the one from s.

It is straightforward to check that μ is also stable under P'. Then we have $\mu'(m) \succeq_m \mu(m)$, which indicates that if m proposes to some woman w in P', then he also proposes to her in P. Now we can prove the lemma by induction on the height of the breath-first search tree on graph T rooted at s. Denote the height of a vertex as $h(v)$. For each vertex with $h(v) = 1$, it must be a woman and has no incoming edge from vertices of M. Therefore, she gets only one proposal from $\mu(w)$ in P. Therefore each man m other than $\mu(w)$ must be matched to a better woman, i.e., $\mu(m) \succ_m w$. Also, as proved above $\mu'(m) \succeq_m \mu(m)$. Then we have $\mu'(m) \succ w$, which means m does not propose to w in P'. The only possible partner for w is $\mu(w)$. Thus, we can conclude that she is matched with $\mu(w)$ in μ', or $\mu'(w) = \mu(w)$.

Assume $\mu'(v) = \mu(v)$ for each v with $h(v) = k$, then for a vertex v' with $h(v') = k + 1$, we prove that we still have $\mu'(v') = \mu(v')$. If $k + 1$ is even, then v' is a man and we consider v''s parent $v = Prt(v')$. From the construction of the graph, there is an edge from v to $\mu(v)$. Henceforth, according to the inductive hypothesis, $\mu'(v) = \mu(v) = v'$, and $\mu(v') = v = \mu'(\mu'(v)) = \mu'(v')$. If $k + 1$ is odd, then v' is a woman and there is an edge pointing to her from v who is the second entry in her received proposal list. On the one hand, each man in $\{m | m \succ_v \mu(v)\}$ is matched with someone who is better than v in μ. As a result, $\mu(m) \succ_m v$. And still $\mu'(m) \succeq_m \mu(m)$, we have $\mu'(m) \succ v$. Therefore m does not propose to her in P'. On the other hand, $\mu(v)$ proposes to v in P' since $\mu(v)$ proposes to her in P. Combining the two sides, we know that $\mu(v)$ is the best man among all those who propose to her. Thus, $\mu'(v) = \mu(v)$. □

PROOF OF THEOREM 4.20. We first construct the suitor graph using μ and compute the corresponding $P(L)$ according to Theorem 4.7. After that, we can compute $Pro(w)$ and $Pro_r(w)$ for each woman w according to $P(L)$. Then we just move the second entry (if exists) of $Pro_r(w)$ to the position right after $\mu(w)$ in each manipulator w's original preference list. Notice that in the modified preference list, no man who is ranked higher than $\mu(w)$ in w's preference list proposes to w, or otherwise the induced matching is unstable with respect to true preference lists. Thus, the orderings of these men is irrelevant to the matching result and we can move $Pro_r(w)$ to the top without affecting the induced matching μ' for the modified lists. According to Lemma B.5, we can conclude that $\mu' = \mu$. □

B.9. Proof of Theorem 4.21

Gonczarowski and Friedgut [2013] consider the sisterhood between manipulators and non-manipulators, and give the following result.

THEOREM B.6 (GONCZAROWSKI AND FRIEDGUT [2013]). *Given agents' strict preferences over agents of the other sex, and a set of manipulators $L \in W$ are allowed to use general manipulations, if no lying woman is worse off, then (1) No woman is worse off; (2) No man is better off.*

Since our setting is a special case of theirs, the above theorem applies to our setting.

PROOF OF THEOREM 4.21. We first prove that all super-strong Nash equilibrium outcomes are Pareto-optimal matchings. Assume on the contrary that a matching μ is induced by a super-strong Nash equilibrium but is not Pareto-optimal. Thus there exists a matching $\mu' \in S_A$ and $\mu' \neq \mu$ such that $\mu'(w) \succeq_w \mu(w), \forall w \in W$ and $\exists w' \in W$ such that $\mu'(w') \succ_{w'} \mu(w')$. It follows that $\mu'(l) = \mu(l), \forall l \in L$ because μ is induced by a super-strong Nash equilibrium, and thus we cannot have $\mu'(l) \succ_l \mu(l)$. Let $\mathcal{R}_\mu$ and $\mathcal{R}_{\mu'}$ be the corresponding set of rotations to μ and μ', respectively. Since $\mu'(w) \succeq_w \mu(w), \forall w \in W$, we have $\mathcal{R}_\mu \subset \mathcal{R}_{\mu'}$ and $\mathcal{R}_{\mu'} \setminus \mathcal{R}_\mu$ is a closed set that still can be eliminated after eliminating $\mathcal{R}_\mu$. Let $R \in Max(\mathcal{R}_{\mu'} \setminus \mathcal{R}_\mu)$. By Lemma B.3, there exists a manipulator l in R. Thus $\mu'(l) \succ \mu(l)$, which contradicts to the fact that $\mu'(l) = \mu(l), \forall l \in L$.

Now we prove that any Pareto-optimal matching can be induced by a super-strong Nash equilibrium. Assume, for purposes of contradiction, that a matching μ is Pareto-optimal but cannot be induced by a super-strong Nash equilibrium. Thus any preference profile that yields matching μ is not a super-strong Nash equilibrium. In particular, we let the manipulators use the inconspicuous manipulation defined above. Let P be the true preference profile and P' be the inconspicuous preference profile that would yield matching $\mu \in S_A(P)$. Since P' is not a super-strong Nash equilibrium, there exists a subset L_s of L, if jointly misreport another preference profile, can make the induced matching to be $\mu' \in S_A(P)$, such that $\forall l \in L_s, \mu'(l) \succeq_l^P \mu(l)$ and $\exists l' \in L_s, \mu'(l') \succ_{l'}^P \mu(l')$. Moreover, there exists $w \in W \setminus L_s$, such that $\mu'(w) \prec_w^P \mu(w)$, since otherwise we have $\mu'(w) \succeq_w^P \mu(w), \forall w \in W$, which contradicts to the Pareto-optimality of μ. Notice that, in inconspicuous preference profile P', we have for all $w \in W$, $m \succ_w^{P'} \mu(w)$ if and only if $m \succ_w^P \mu(w)$ since the order of men ranked higher than $\mu(w)$ in $P'(w)$ is exactly the same as $P(w)$. Therefore, for all $l \in L_s$, $\mu'(l) \succ_l^{P'} \mu(l)$. However, according to Theorem B.6, since no manipulators are worse off according to P', we have that no women are worse off according to P'. Also, for all $w \in W \setminus L_s$, $\mu'(w) \succeq_l^{P'} \mu(w)$ implies $\mu'(w) \succeq_l^P \mu(w)$. Thus, $\mu' \succeq_W^P \mu$. A contradiction. □

B.10. Proof of Theorem 4.22

Clearly, this problem is in NP since given a preference profile, we can apply Gale-Shapley algorithm to generate the induced matching and verify the solution. In order to show the NP-completeness, we reduce 3-SAT to this problem. Given an instance of 3-SAT ϕ, suppose the variable set is $V = \{x_1, \ldots, x_n\}$, the corresponding literal set is $L = \{+x_i, -x_i \mid 1 \leq i \leq n\}$, and the clause set is $\{c_1, \ldots, c_m\}$, where $c_j = (l_j^1, l_j^2, l_j^3)$. We construct an instance of our problem $G(\phi)$ with $N = 6n + 2m$ and

$$\begin{aligned}
M = &\{m_{x_i}^{+1}, m_{x_i}^{+2}, m_{x_i}^{+3} \mid \forall 1 \leq i \leq n\} \cup \{m_{x_i}^{-1}, m_{x_i}^{-2}, m_{x_i}^{-3} \mid \forall 1 \leq i \leq n\} \\
&\cup \{m_{c_j}^{l} \mid \forall 1 \leq j \leq m\} \cup \{m_{c_j}^{r} \mid \forall 1 \leq j \leq m\} \\
W = &\{w_{x_i}^{+1}, w_{x_i}^{+2}, w_{x_i}^{+3} \mid \forall 1 \leq i \leq n\} \cup \{w_{x_i}^{-1}, w_{x_i}^{-2}, w_{x_i}^{-3} \mid \forall 1 \leq i \leq n\} \\
&\cup \{w_{c_j}^{l} \mid \forall 1 \leq j \leq m\} \cup \{w_{c_j}^{r} \mid \forall 1 \leq j \leq m\}
\end{aligned}$$

The set of manipulators is

$$L = \{w_{x_i}^{+_2} \mid \forall 1 \leq i \leq n\} \cup \{w_{x_i}^{-_2} \mid \forall 1 \leq i \leq n\} \cup \{w_{c_j}^{r} \mid \forall 1 \leq j \leq m\}$$

The preference lists of each agent is specified as follows (the "$\cdots$" part at the end can be anything). For all $1 \leq i \leq n$ and each x_i, in the positive side (with superscript $+$),

$$\begin{aligned}
P(m_{x_i}^{+_1}) &= w_{x_i}^{+_1} \succ w_{x_i}^{+_2} \succ w_{x_i}^{-_3} \succ \cdots \\
P(m_{x_i}^{+_2}) &= w_{x_i}^{+_2} \succ w_{x_i}^{+_1} \succ \cdots \\
P(w_{x_i}^{+_1}) &= m_{x_i}^{+_2} \succ m_{x_i}^{+_1} \succ \cdots \\
P(w_{x_i}^{+_2}) &= m_{x_i}^{-_3} \succ m_{x_i}^{+_1} \succ m_{x_i}^{+_2} \succ m_{x_i}^{+_3} \succ \cdots \\
P(w_{x_i}^{+_3}) &= m_{x_i}^{-_1} \succ m_{x_i}^{+_3} \succ \cdots
\end{aligned}$$

In the negative side (with superscript $-$), similarly,

$$\begin{aligned}
P(m_{x_i}^{-_1}) &= w_{x_i}^{-_1} \succ w_{x_i}^{-_2} \succ w_{x_i}^{+_3} \succ \cdots \\
P(m_{x_i}^{-_2}) &= w_{x_i}^{-_2} \succ w_{x_i}^{-_1} \succ \cdots \\
P(w_{x_i}^{-_1}) &= m_{x_i}^{-_2} \succ m_{x_i}^{-_1} \succ \cdots \\
P(w_{x_i}^{-_2}) &= m_{x_i}^{+_3} \succ m_{x_i}^{-_1} \succ m_{x_i}^{-_2} \succ m_{x_i}^{-_3} \succ \cdots \\
P(w_{x_i}^{-_3}) &= m_{x_i}^{+_1} \succ m_{x_i}^{-_3} \succ \cdots
\end{aligned}$$

Suppose $+x_i \in c_{k_j}$ for all $1 \leq j \leq K_i^+$. The preference list of $m_{x_i}^{+_3}$ is

$$P(m_{x_i}^{+_3}) = w_{x_i}^{+_2} \succ w_{x_i}^{+_3} \succ w_{c_{k_1}}^{l} \succ w_{c_{k_2}}^{l} \succ \cdots \succ w_{c_{k_{K_i^+}}}^{l} \succ w_{x_i}^{-_2} \succ \cdots$$

Similarly, Suppose $-x_i \in c_{k_j}$ for all $1 \leq j \leq K_i^-$. The preference list of $m_{x_i}^{r_3}$ is

$$P(m_{x_i}^{-_3}) = w_{x_i}^{-_2} \succ w_{x_i}^{-_3} \succ w_{c_{k_1}}^{l} \succ w_{c_{k_2}}^{l} \succ \cdots \succ w_{c_{k_{K_i^-}}}^{l} \succ w_{x_i}^{+_2} \succ \cdots$$

Finally, we specify the preference lists for the agent with subscript c_j. For all $1 \leq j \leq m$,

$$\begin{aligned}
P(m_{c_j}^{l}) &= w_{c_j}^{l} \succ w_{c_j}^{r} \succ \cdots \\
P(m_{c_j}^{r}) &= w_{c_j}^{r} \succ w_{c_j}^{l} \succ \cdots \\
P(w_{c_j}^{r}) &= m_{c_j}^{l} \succ m_{c_j}^{r} \succ \cdots
\end{aligned}$$

Suppose $c_j = (s^1\ x_{j_1}) \vee (s^2\ x_{j_2}) \vee (s^3\ x_{i_3})$. where $s^1, s^2, s^3 \in \{-, +\}$. The preference list of $w_{c_j}^{l}$ is [4]

$$P(w_{c_j}^{l}) = m_{c_j}^{r} \succ m_{x_{j_1}}^{s_3^1} \succ m_{x_{j_2}}^{s_3^2} \succ m_{x_{j_3}}^{s_3^3} \succ m_{c_j}^{l} \succ \cdots$$

To complete the reduction, we prove that ϕ is satisfiable if and only if $G(\phi)$ has a solution, i.e., there exists a strategy profile, whose induced matching is stable and strictly better off for all manipulators.

First, notice the stable matching μ generated by true preference lists is $\mu(m_{x_i}^{+_k}) = w_{x_i}^{+_k}$, $\mu(m_{x_i}^{-_k}) = w_{x_i}^{-_k}$ for all $1 \leq i \leq n$, $1 \leq k \leq 3$ and $\mu(m_{c_j}^{l}) = w_{c_j}^{l}$, $\mu(m_{c_j}^{r}) = w_{c_j}^{r}$ for all $1 \leq j \leq m$. Before providing proofs for both directions, we prove following lemmas first to establish intuitions.

[4] If $s^k = +$, then $s_3^k = +_3$; otherwise, if $s^k = -$, $s_3^k = -_3$.

LEMMA B.7. *For all $i \in [n]$, woman $w_{x_i}^{+_2}$ can perform single-agent manipulation to be matched with $m_{x_i}^{+_1}$.*

PROOF. $w_{x_i}^{+_2}$ can manipulate her preference list to $P(w_{x_i}^{+_2}) = m_{x_i}^{-_3} \succ m_{x_i}^{+_1} \succ m_{x_i}^{+_3} \succ m_{x_i}^{+_2} \succ \cdots$. □

LEMMA B.8. *For all $i \in [n]$, woman $w_{x_i}^{+_2}$ can perform single-agent manipulation to be matched with $m_{x_i}^{-_3}$.*

PROOF. $w_{x_i}^{+_2}$ can manipulate her preference list to $P(w_{x_i}^{+_2}) = m_{x_i}^{-_3} \succ m_{x_i}^{+_3} \succ m_{x_i}^{+_1} \succ m_{x_i}^{+_2} \succ \cdots$. □

By symmetry of construction, we have for each $1 \leq i \leq n$, woman $w_{x_i}^{-_2}$ can perform single-agent manipulation to be matched with $m_{x_i}^{-_1}$ or $m_{x_i}^{+_3}$.

LEMMA B.9. *$w_{x_i}^{+_2}$ and $w_{x_i}^{-_2}$ cannot manipulate to be matched with $m_{x_i}^{-_3}$ and $m_{x_i}^{+_3}$ respectively at the same time, in any feasible permutation manipulation, while it is possible for them to manipulate to be matched with $m_{x_i}^{+_1}$ and $m_{x_i}^{-_1}$, $m_{x_i}^{-_3}$ and $m_{x_i}^{-_1}$, or, $m_{x_i}^{+_1}$ and $m_{x_i}^{+_3}$, respectively.*

Before proving Lemma B.9, we first prove the following lemma,

LEMMA B.10. *If a the induced matching of a permutation manipulation on $G(\phi)$ is stable with respect to true preference lists, then*

(*1*) *For all $i \in [n]$, $m_{x_i}^{s_3}$, he cannot make proposals to any woman ranked below $w_{x_i}^{-s_2}$ in his true preference list; Moreover, he cannot be matched with any $w_{c_j}^{l}$;*
(*2*) *For all $i \in [n]$, $m_{x_i}^{s_1}$ and $m_{x_i}^{s_1}$ with $s \in \{+, -\}$, he can only make proposals to woman $w_{x_i}^{s'_k}$ with $s' \in \{+, -\}$ and $k \in \{1, 2, 3\}$;*
(*3*) *For all $j \in [m]$, both $m_{c_j}^{l}$ and $m_{c_j}^{r}$, he can only make proposals to $w_{c_j}^{l}$ and $w_{c_j}^{r}$;*

PROOF. Denote

$$W_i = \{w_{x_i}^{+_1}, w_{x_i}^{+_2}, w_{x_i}^{+_3}, w_{x_i}^{-_1}, w_{x_i}^{-_2}, w_{x_i}^{-_3}\}$$

and

$$M_i = \{m_{x_i}^{+_1}, m_{x_i}^{+_2}, m_{x_i}^{+_3}, m_{x_i}^{-_1}, m_{x_i}^{-_2}, m_{x_i}^{-_3}\}.$$

First, for $m_{x_i}^{s_3}$ with $j \neq i$, $s \in \{+, -\}$, since $w_{x_i}^{-s_2}$ puts $m_{x_j}^{s_3}$ as the favorite candidate, if $m_{x_i}^{s_3}$ proposes to any woman ranked below $w_{x_i}^{-s_2}$ in his true preference list, the induced matching is unstable with respect to true preference lists. Moreover, if $m_{x_i}^{s_3}$ proposes to some $w_{c_j}^{l}$, then $w_{c_j}^{l}$ accepts $m_{x_{j_1}}^{s_3}$ and rejects $m_{c_j}^{l}$, next, $w_{c_j}^{r}$ accepts $m_{c_j}^{l}$ and rejects $m_{c_j}^{r}$, and finally, $w_{c_j}^{l}$ accepts $m_{c_j}^{r}$ and rejects $m_{x_{j_1}}^{s_3}$.

Second, except $m_{x_i}^{+_3}$ and $m_{x_i}^{-_3}$, all men in M_i only propose to women in W_i before they propose to the woman ranking him as the highest. Therefore, with similar arguments, we conclude that $m_{x_i}^{s_1}$ and $m_{x_i}^{s_1}$ with $s \in \{+, -\}$, he can only make proposals to woman $w_{x_i}^{s'_k}$ with $s' \in \{+, -\}$ and $k \in \{1, 2, 3\}$

Third, since $w_{c_j}^{l}$ ranks $m_{c_j}^{r}$ as favorite and $w_{c_j}^{r}$ ranks $m_{c_j}^{l}$ as favorite, according to the preference lists of $m_{c_j}^{l}$ and $m_{c_j}^{r}$, we can conclude they can only make proposals to $w_{c_j}^{l}$ and $w_{c_j}^{r}$; □

PROOF OF LEMMA B.9. To achieve other combinations, $w_{x_i}^{+_2}$ and $w_{x_i}^{-_2}$ can manipulate their preference lists by following the manipulation in Lemma B.7 and Lemma B.8 according to their target partners.

We prove the remaining case by contradiction. Suppose $w^{+_2}_{x_i}$ and $w^{-_2}_{x_i}$ can manipulate to a matching μ such that they are matched with $m^{-_3}_{x_i}$ and $m^{+_3}_{x_i}$. Then, since $w^{+_2}_{x_i}$ is matched with $m^{-_3}_{x_i}$, the closed set of M-rotation

$$(\{m^{+_1}_{x_i}, m^{-_3}_{x_i}\}, \{w^{+_2}_{x_i}, w^{-_3}_{x_i}\}, \{w^{-_3}_{x_i}, w^{+_2}_{x_i}\})$$

must be eliminated, which contains M-rotation

$$(\{m^{+_2}_{x_i}, m^{+_1}_{x_i}\}, \{w^{+_2}_{x_i}, w^{+_1}_{x_i}\}, \{w^{+_1}_{x_i}, w^{+_2}_{x_i}\}).$$

Similarly, since $w^{-_2}_{x_i}$ is matched with $m^{+_3}_{x_i}$, the closed set of M-rotation

$$(\{m^{-_1}_{x_i}, m^{+_3}_{x_i}\}, \{w^{-_2}_{x_i}, w^{+_3}_{x_i}\}, \{w^{+_3}_{x_i}, w^{-_2}_{x_i}\})$$

must be eliminated, which contains M-rotation

$$(\{m^{-_2}_{x_i}, m^{-_1}_{x_i}\}, \{w^{-_2}_{x_i}, w^{-_1}_{x_i}\}, \{w^{-_1}_{x_i}, w^{-_2}_{x_i}\}).$$

Therefore, all of $W_i = \{w^{+_1}_{x_i}, w^{+_2}_{x_i}, w^{+_3}_{x_i}, w^{-_1}_{x_i}, w^{-_2}_{x_i}, w^{-_3}_{x_i}\}$ have received more than one proposals. Moreover, according to Lemma B.10, they are matched with one of $M_i = \{m^{+_1}_{x_i}, m^{+_2}_{x_i}, m^{+_3}_{x_i}, m^{-_1}_{x_i}, m^{-_2}_{x_i}, m^{-_3}_{x_i}\}$.

Henceforth, $\mu \in S_A$ only if there is some man $m \notin M_i$ having made proposal to some $w \in W_i$ in order to create connections from s. However, according to Lemma B.10, if $\mu \in S_A$, no other man $m \notin M_i$ can make proposal to some $w \in W_i$. □

We point out that in this lemma, our construction contains the example in Table II as a gadget. According to this lemma, given an outcome of manipulation, we construct the assignment as follows. $+x_i$ is assigned *true* if and only if $w^{+_2}_{x_i}$ is matched with $m^{-_3}_{x_i}$; otherwise, $-x_i$ is assigned *true*. Next lemma guarantees that such assignment is a satisfiable assignment for ϕ.

LEMMA B.11. *For all $j \in [m]$, suppose $c_j = (s^1\ x_{j_1}) \vee (s^2\ x_{j_2}) \vee (s^3\ x_{j_3})$. Then, after manipulation, woman $w^r_{c_j}$ can be better off if and only if at least one $w^{s^k_2}_{x_{j_k}}$ is matched with $m^{-s^k_3}_{x_{j_k}}$ for $k \in \{1,2,3\}$.*

PROOF. *if* direction: Without loss of generality, suppose $w^{s_2}_{x_{j_1}}$ with $s = s^1$ is matched with $m^{-s_3}_{x_{j_1}}$ and thus, $m^{-s_3}_{x_{j_1}}$ has made proposal to $w^{-s_2}_{x_{j_1}}, w^{-s_3}_{x_{j_1}} \succ \cdots \succ w^l_{c_j} \succ \cdots \succ w^{s_2}_{x_{j_1}} \succ \cdots$. Thus, $w^l_{c_j}$ accepts $m^{-s_3}_{x_{j_1}}$ and rejects $m^l_{c_j}$, next, $w^r_{c_j}$ accepts $m^l_{c_j}$ and rejects $m^r_{c_j}$, and finally, $w^l_{c_j}$ accepts $m^r_{c_j}$ and rejects $m^{-s_3}_{x_{j_1}}$. Therefore, $w^r_{c_j}$ is better off. Moreover, if more than one $w^{s^k_2}_{x_{j_k}}$ is matched with $m^{-s^k_3}_{x_{j_k}}$, it does not change the matching of $w^r_{c_j}$ since she is already matched with her favorite one.

only if direction: If no $w^{s^k_2}_{x_{j_k}}$ is matched with $m^{-s^k_3}_{x_{j_k}}$, notice that no $m^{-s^k_3}_{x_{j_k}}$ makes proposal to $w^l_{c_j}$ since from argument in "*if direction*", we can see that if $m^{-s^k_3}_{x_{j_k}}$ makes proposal to $w^l_{c_j}$, $w^{s^k_2}_{x_{j_k}}$ is matched with $m^{-s^k_3}_{x_{j_k}}$. Therefore, if $w^r_{c_j}$ is better off, then $w^r_{c_j}$ is matched with $m^l_{c_j}$ and $w^l_{c_j}$ is matched with $m^r_{c_j}$, and notice that, $w^l_{c_j}, w^r_{c_j}$ have received more than one proposals. Henceforth, the matching after manipulation is in S_A only if there is some man outside $m^l_{c_j}, m^r_{c_j}$ having made proposal to one of $w^l_{c_j}, w^r_{c_j}$ in order to create an edge pointing to the strongly connected component. However, according to Lemma B.10, we can conclude that no man outside $m^l_{c_j}, m^r_{c_j}$ having made proposal to one of $w^l_{c_j}, w^r_{c_j}$. □

With Lemma B.7, Lemma B.8, Lemma B.9 and Lemma B.11, we are ready to complete our reduction by showing ϕ is satisfiable if and only if $G(\phi)$ has a solution.

LEMMA B.12. *ϕ is satisfiable only if $G(\phi)$ has a solution.*

PROOF. Suppose $(l'_1, \ldots, l'_n)$ is a satisfiable assignment. For all $i \in [n]$,

(1) if $l'_i = +x_i$: $w_{x_i}^{+2}$ manipulates to $m_{x_i}^{-3}$ and $w_{x_i}^{-2}$ manipulates to $m_{x_i}^{-1}$;
(2) if $l'_i = -x_i$: $w_{x_i}^{+2}$ manipulates to $m_{x_i}^{+1}$ and $w_{x_i}^{-2}$ manipulates to $m_{x_i}^{+3}$;

According to Lemma B.9, the matching induced by this manipulation is in S_A. Moreover, since $(l'_1, \ldots, l'_n)$ is a satisfiable assignment, from Lemma B.11, for all $j \in [m]$, $w_{c_j}^{r}$ is better off. □

LEMMA B.13. *ϕ is satisfiable if $G(\phi)$ has a solution.*

PROOF. From Lemma B.9, for each $1 \leq i \leq n$, $w_{x_i}^{+2}$ and $w_{x_i}^{-2}$ cannot manipulate to be matched with $m_{x_i}^{-3}$ and $m_{x_i}^{+3}$ respectively. Therefore, we create the assignment as follows:

(1) $+x_i$ is assigned *true* if and only if $w_{x_i}^{+2}$ is matched with $m_{x_i}^{-3}$;
(2) otherwise, $-x_i$ is assigned *true*;

Moreover, from Lemma B.11, since for all $1 \leq j \leq m$ with $c_j = (s^1\, x_{j_1}) \vee (s^2\, x_{j_2}) \vee (s^3\, x_{j_3})$, $w_{c_j}^{r}$ is better off, at least one $w_{x_{j_k}}^{s_2^k}$ is matched with $m_{x_{j_k}}^{-s_3^k}$ for $k \in \{1, 2, 3\}$. Thus, the assignment we create must be a satisfiable assignment for ϕ. □

B.11. Proof of Theorem 4.23

LEMMA B.14. *In our construction, if all manipulators are better off in a matching, the matching must be stable.*

PROOF. First, we point out that if a manipulation induces an unstable matching, then some woman must reject the best proposal she could have in the entire process. Henceforth, she must be a manipulator, while $L = \{w_{x_i}^{+2} \mid \forall 1 \leq i \leq n\} \cup \{w_{x_i}^{-2} \mid \forall 1 \leq i \leq n\} \cup \{w_{c_j}^{r} \mid \forall 1 \leq j \leq m\}$ in our construction.

Notice that for all $1 \leq j \leq m$, $w_{c_j}^{r}$ can only be matched with $m_{c_j}^{l}$ if she is better off, and thus she cannot reject her best received proposal.

The remaining manipulators are $w_{x_i}^{+2}$ and $w_{x_i}^{-2}$ for $1 \leq i \leq n$. Consider $w_{x_i}^{+2}$ and the argument for $w_{x_i}^{-2}$ is similar due to symmetry of construction. Since $w_{x_i}^{+2}$ is better off, $w_{x_i}^{+2}$ must be matched with either $m_{x_i}^{-3}$ or $m_{x_i}^{+1}$. In the case that she rejects her best received proposal, $w_{x_i}^{+2}$ must be matched with $m_{x_i}^{+1}$ and reject $m_{x_i}^{-3}$. However, if $w_{x_i}^{+2}$ is matched with $m_{x_i}^{+1}$, $m_{x_i}^{+1}$ stops proposing after meeting $w_{x_i}^{+2}$, and thus, $w_{x_i}^{-3}$ cannot reject $m_{x_i}^{-3}$ since $w_{x_i}^{-3}$ is a non-manipulator and she does not receive her favorite man $m_{x_i}^{+1}$ to reject her second favorite man $m_{x_i}^{-3}$. Therefore, $m_{x_i}^{-3}$ has no chance to propose to $w_{x_i}^{+2}$ and get rejected. □

Combining Theorem 4.22 and Lemma B.14, we can conclude our theorem.

B.12. Proof of Theorem 4.24

PROOF OF THEOREM 4.24. Since computing the number of satisfiable assignment for 3-SAT problem is #P-complete, we only need to show that our reduction is *parsimonious*, i.e., the numbers of solutions in each problem are the same.

First, we show that given one satisfiable assignment for 3-SAT problem, we can construct a solution in PARETO-BETTER. According to Lemma B.12, we can construct a solution that makes all manipulators better off. Thus, it is sufficient to show that the constructed solution is also Pareto-optimal. In fact, for all $1 \leq i \leq n$, either $w_{x_i}^{+2}$ or $w_{x_i}^{+2}$ is matched with her favorite partner, but it is impossible for them to be matched with their favorite partners simultaneously. Moreover, for all $1 \leq j \leq m$, $w^r_{c_j}$ is matched with her favorite partner. Thus, such a solution must be Pareto-optimal.

Second, we show that given a solution to PARETO-BETTER, we can construct a satisfiable assignment for 3-SAT problem. From Lemma B.13, we have shown that given a matching that makes all manipulators better off, we can construct a satisfiable assignment. Thus, given a solution to PARETO-BETTER, we are also able to construct a satisfiable assignment for 3-SAT problem. □

C. EXAMPLES

C.1. Example of general manipulation to an unstable matching

Consider one manipulator w to keep her W-pessimal partner m and reject any better proposals. Therefore, it is equivalent to a manipulation game by removing w from W, m from M and it is possible that in the remaining manipulation game, the W-optimal matching is weakly better off than the W-optimal matching in the original game, though it is unstable with respect to true preference lists. Thus, with the help of manipulator w, a coalition can have a further manipulation to make everyone weakly better off and at least one strictly better off (see Table IV).

Table IV. Example of general manipulation to unstable matching

Men's preference lists

m_1	w_1	w_3	w_2
m_2	w_2	w_1	$-$
m_3	w_3	$-$	$-$

Women's preference lists

w_1	m_2	m_1	$-$
w_2	m_1	m_2	$-$
w_3	m_1	m_3	$-$

The only stable matching under true preference lists is $\{(m_1, w_1), (m_2, w_2), (m_3, w_3)\}$. However, consider following manipulation (see Table V).

Table V. Manipulation of general manipulation to unstable matching

Men's preference lists

m_1	w_1	w_3	w_2
m_2	w_2	w_1	$-$
m_3	w_3	$-$	$-$

Women's preference lists

w_1	m_2	$-$	$-$
w_2	m_1	$-$	$-$
w_3	m_3	$-$	$-$

After manipulation, the only stable matching is $\{(m_1, w_2), (m_2, w_1), (m_3, w_3)\}$, in which w_1 and w_2 are strictly better off while w_3 remains the same. However, this matching is unstable with respect to true preference lists.

C.2. Example of permutation manipulation to an unstable matching

The following example demonstrates the output of Algorithm 1 is no longer a super-strong Nash equilibrium, without the feasibility assumption (see Table VI). Thus, the feasibility assumption is essential for our results to hold.

Table VI. Example of permutation manipulation to unstable matching

Men's preference lists

m_1	w_1	w_3	w_4	w_2
m_2	w_2	w_3	w_1	w_4
m_3	w_3	w_1	w_2	w_4
m_4	w_3	w_4	w_1	w_2

Women's preference lists

w_1	m_3	m_1	m_2	m_4
w_2	m_3	m_2	m_1	m_4
w_3	m_2	m_3	m_4	m_1
w_4	m_1	m_2	m_3	m_4

The stable matching is $\{(m_1, w_1), (m_2, w_2), (m_3, w_3), (m_4, w_4)\}$, while only w_3 receives more than one proposals in the entire process. Therefore, suppose the manipulators are w_1 and w_3, in order to manipulate, the only way is that w_3 keeps m_4 and rejects m_3, and m_3 proposes to w_1.

Table VII. Process after manipulation

Women	w_1	w_2	w_3	w_4
Round 1	m_1	m_2	m_4	$-$
Round 2	m_1, m_3	m_2	m_4	$-$
Round 3	m_1	m_3, m_2	m_4	$-$
Round 4	m_1	m_3	m_2, m_4	$-$
Round 5	m_1	m_3	m_2	m_4

Notice that, if w_1 keeps her favorite one m_3 and rejects m_1, then in the next round, m_1 would propose to w_3 and no matter which one w_3 rejects, either m_1 or m_4 would propose to w_4 and end the process, leaving w_3 worse off. Therefore, the stable matching is Pareto-optimal under feasibility assumption. In order to go beyond Pareto-optimal matching under feasibility assumption, we resort to unstable matching, i.e., w_1 rejects m_3 and still keeps m_1. (see the entire process in Table VII)

Since w_1 rejects her best received proposal, the induced matching is unstable. However, w_1, w_4 are matched with the same man as in the original matching while w_2 and w_3 are better off.

D. NUMBER OF PARETO-OPTIMAL MATCHINGS

PROPOSITION D.1. *There are exactly 2^n different induced matchings, which is Pareto-optimal and weakly better off for all manipulators, in $G(\phi)$.*

PROOF. Notice that, for all $1 \leq j \leq m$, the partner of $w_{c_j}^r$ is determined by whether there is a man other than $m_{c_j}^l$ who makes a proposal to $w_{c_j}^l$, which is determined by the matching between $w_{x_i}^{s_2}$ and $w_{x_i}^{-s_3}$, for $1 \leq i \leq n$ and $s \in \{+, -\}$. Therefore, we can count the total number according to the number of different matching for $w_{x_i}^{s_2}$. From Lemma B.9, under Pareto-optimal constraints, there are 2 different choices for each pair of $w_{x_i}^{+_2}$ and $w_{x_i}^{-_2}$, and they are independent. Thus, the total number of solutions is exactly 2^n. □