

# The Maximum-weight Stable Matching Problem: Duality and Efficiency

Xujin Chen<sup>a\*</sup> Guoli Ding<sup>b†</sup> Xiaodong Hu<sup>a‡</sup> Wenan Zang<sup>c§</sup>

<sup>a</sup> Institute of Applied Mathematics, Chinese Academy of Sciences  
Beijing 100190, China

<sup>b</sup> Mathematics Department, Louisiana State University  
Baton Rouge, LA 70803, USA

<sup>c</sup> Department of Mathematics, The University of Hong Kong  
Hong Kong, China

## Abstract

Given a preference system  $(G, \prec)$  and an integral weight function defined on the edge set of  $G$  (not necessarily bipartite), the maximum-weight stable matching problem is to find a stable matching of  $(G, \prec)$  with maximum total weight. In this paper we study this  $NP$ -hard problem using linear programming and polyhedral approaches. We show that the Rothblum system for defining the fractional stable matching polytope of  $(G, \prec)$  is totally dual integral if and only if this polytope is integral if and only if  $(G, \prec)$  has a bipartite representation. We also present a combinatorial polynomial-time algorithm for the maximum-weight stable matching problem and its dual on any preference system with a bipartite representation. Our results generalize Király and Pap's theorem on the maximum-weight stable-marriage problem and rely heavily on their work.

**MSC 2000 subject classification.** Primary: 90C10, 90C27, 90C57.

**OR/MS subject classification.** Primary: Programming/graphs.

**Key words.** Stable matching, linear system, integral polytope, total dual integrality, polynomial-time algorithm.

---

\*Supported in part by NSF of China under Grant 10771209 and Chinese Academy of Sciences under Grant kjcx-yw-s7.

†Supported in part by NSA grant H98230-10-1-0186 and NSF grant DMS-1001230.

‡Supported in part by NSF of China under Grant 10721101, 10771209 and 10928102 and Chinese Academy of Sciences under Grant kjcx-yw-s7.

§Supported in part by the Research Grants Council of Hong Kong, the Overseas and Hong Kong, Macau Young Scholars Collaborative Research Fund of the National Science Foundation of China, and Seed Funding for Basic Research of HKU. Corresponding author. E-mail: wzang@maths.hku.hk.

## 1 Introduction

Let  $G = (V, E)$  be a graph. For each  $v \in V$ , let  $\delta(v)$  be the set of all edges incident with  $v$  and let  $\prec_v$  be a strict linear order on  $\delta(v)$ . We call  $\prec_v$  the *preference* of  $v$  and say that  $v$  *prefers*  $e$  to  $f$  if  $e \prec_v f$ . Let  $\prec$  be the collection of all these  $\prec_v$  for  $v \in V$ . We call the pair  $(G, \prec)$  a *preference system*. In particular,  $(G, \prec)$  is referred to as a *bipartite preference system* if  $G$  is a bipartite graph. For convenience, we write  $e \preceq_v f$  and say that  $e$  *dominates*  $f$  at vertex  $v$  if  $e \prec_v f$  or  $e = f$ . An edge  $e$  is said to *dominate* an edge  $f$  if they have a common end  $v$  such that  $e \preceq_v f$ . Let  $M$  be a matching of  $G$ . We call  $M$  *stable* if each edge of  $G$  is dominated by some edge in  $M$ . The *stable matching problem* (SMP) is to determine if  $G$  contains a stable matching. The origin of this problem can be traced back to 1962 when Gale and Shapley [4] proposed the well-known stable marriage problem (which corresponds to the case where  $G$  is a bipartite graph); since then the SMP and its variants have been subjects of extensive research, see, for instance, the books by Knuth [9], Gusfield and Irving [6], Roth and Sotomayor [10], and the references therein. Gale and Shapley [4] proved that, for the stable marriage problem, a stable matching always exists. Irving [7] devised the first polynomial-time algorithm for finding a stable matching of  $(G, \prec)$ , if any, or declaring that no such matching exists, where  $G$  is not necessarily bipartite. With an attempt to improve Irving's algorithm [7], Tan [14] obtained a necessary and sufficient condition for the existence of a (perfect) stable matching in terms of forbidden structures, the so-called stable partitions with odd parties.

The present paper is devoted to the *maximum-weight stable matching problem* (MWSMP), which takes a preference system  $(G, \prec)$  and an integral weight function  $w$  defined on  $E$  as input, and aims to find a stable matching of  $(G, \prec)$  with maximum total weight as output, where  $G = (V, E)$ . As shown by Feder [3], this problem is *NP*-hard in general, so there is no polynomial-time algorithm for solving it exactly unless  $NP = P$ . In this paper we study the MWSMP using linear programming and polyhedral approaches, which were first applied to stable matchings by Vande Vate [16] and Rothblum [12]. It was discovered by Rothblum [12] that the convex hull of stable matchings of a bipartite preference system can be described by a very simple system of linear inequalities; this polyhedral description was later extended and further developed in different aspects by Roth, Rothblum and Vande Vate [11], Abeledo and Rothblum [1], Teo and Sethuraman [15], and Király and Pap [8]. The objective of this paper is to characterize all preference systems for which the above-mentioned Rothblum system is totally dual integral and to present a combinatorial polynomial-time algorithm for the MWSMP and its dual on these preference systems.

We introduce some notations and terminology before proceeding. As usual, let  $\mathbb{R}$  (resp.  $\mathbb{R}_+$ ) be the set of all (resp. nonnegative) real numbers, and let  $\mathbb{Z}$  (resp.  $\mathbb{Z}_+$ ) be the set of all (resp. nonnegative) integers. A subset  $P$  of  $\mathbb{R}^n$  is called a *polytope* if it is the convex hull of finitely many vectors in  $\mathbb{R}^n$ . A point  $x$  in  $P$  is called a *vertex* or an *extreme point* if there exist no distinct points  $y$  and  $z$  in  $P$  and  $\alpha \in (0, 1)$  such that  $x = \alpha y + (1 - \alpha)z$ . It is well known that  $P$

is actually the convex hull of its vertices, and that there exists a linear system  $Ax \leq b$  such that  $P = \{x : Ax \leq b\}$ . We call  $P$  *integral* if each of its vertices is an integral vector. By a theorem in mathematical programming,  $P$  is integral if and only if the maximum in the LP-duality equation

$$\max\{w^T x : Ax \leq b\} = \min\{y^T b : y^T A = w^T, y \geq 0\} \quad (1.1)$$

has an integral optimal solution, for every integral vector  $w$  for which the optimum is finite. If, instead, the minimum in the equation enjoys this property, then the system  $Ax \leq b$  is called *totally dual integral* (TDI). The model of TDI systems plays a crucial role in combinatorial optimization, and serves as a general framework for establishing various min-max theorems because, as shown by Edmonds and Giles [2], total dual integrality implies primal integrality: if  $Ax \leq b$  is TDI and  $b$  is integral, then  $P$  is integral.

Let  $(G, \prec)$  be a preference system, where  $G = (V, E)$ . For each  $e \in E$ , let  $\varphi(e)$  denote the set of all edges of  $G$  that dominate  $e$ , and let  $\psi(e)$  denote the set of all edges of  $G$  that are dominated by  $e$ . For each  $c \in \mathbb{R}^E$  and  $S \subseteq E$ , let  $c(S) = \sum_{s \in S} c(s)$ . As observed by Abeledo and Rothblum [1], the incidence vectors of stable matchings of  $(G, \prec)$  are precisely integral solutions  $x \in \mathbb{Z}^E$  of the following inequalities:

$$x(\delta(v)) \leq 1 \quad \forall v \in V, \quad (1.2)$$

$$x(\varphi(e)) \geq 1 \quad \forall e \in E, \quad (1.3)$$

$$x(e) \geq 0 \quad \forall e \in E. \quad (1.4)$$

Let  $SM(G, \prec)$  denote the convex hull of incidence vectors of all stable matchings of  $(G, \prec)$ , and let  $FSM(G, \prec)$  denote the set of all vectors  $x \in \mathbb{R}^E$  satisfying (1.2) – (1.4). Clearly,  $SM(G, \prec) \subseteq FSM(G, \prec)$ . In the literature,  $SM(G, \prec)$  and  $FSM(G, \prec)$  are called the *stable matching polytope* and *fractional stable matching polytope* of  $(G, \prec)$ , respectively.

For each  $x \in FSM(G, \prec)$ , set

- $E_+(x) = \{e \in E : x_e > 0\}$ ,
- $E_\alpha(x) = \{e \in E : x_e = \alpha\}$  for any  $\alpha \in \mathbb{R}$ , and
- $T(x) = \{uv \in E : \exists e, f \in E_+(x) \text{ such that } uv \preceq_u e \text{ and } uv \preceq_v f\}$ .

Notice that  $T(x)$  consists of every edge  $uv$  of  $G$  that dominates some edges in  $E_+(x)$  at both  $u$  and  $v$ . Let  $G_{T(x)}$  denote the subgraph of  $G$  induced by all edges in  $T(x)$ . As we shall see,  $G_{T(x)}$  can be used to characterize all vertices  $x$  of  $FSM(G, \prec)$ . A cycle  $C = v_1 v_2 \dots v_k v_1$  of  $G$  is said to have a *cyclic preference* in  $(G, \prec)$  if  $v_{i-1} v_i \preceq_{v_i} v_i v_{i+1}$  for all  $1 \leq i \leq k$  or  $v_{i+1} v_i \preceq_{v_i} v_i v_{i-1}$  for all  $1 \leq i \leq k$ , where  $v_{k+1} = v_1$ . Moreover, a vector  $y$  is called *half-integral* if all coordinates of  $2y$  are integral.

Let  $\pi(G, \prec)$  stand for the system of linear inequalities (1.2)–(1.4). Rothblum [12] proved that the stable matching polytope of a bipartite preference system  $(G, \prec)$  is completely determined by  $\pi(G, \prec)$ .

**Theorem 1.1** (Rothblum [12]) *Let  $(G, \prec)$  be a bipartite preference system. Then*

$$SM(G, \prec) = FSM(G, \prec).$$

For an arbitrary preference system, Abeledo and Rothblum [1] showed that the vertices of its fractional stable matching polytope are always half-integral; they also gave a description of all vertices of this polytope.

**Theorem 1.2** (Abeledo-Rothblum [1]) *Let  $(G, \prec)$  be a preference system and let  $x$  be a point in  $FSM(G, \prec)$ . Then  $x$  is a vertex of  $FSM(G, \prec)$  if and only if*

- (i)  $x$  is half-integral and
- (ii) each component of  $G_{T(x)}$  having edges in  $E_{1/2}(x)$  contains at least one odd cycle.

In [8], Király and Pap obtained the following strengthening of Theorem 1.1.

**Theorem 1.3** (Király-Pap [8]) *Let  $(G, \prec)$  be a bipartite preference system. Then  $\pi(G, \prec)$  is totally dual integral.*

Let  $(G, \prec)$  be a preference system with  $G = (V, E)$ . Recall that for each  $v \in V$ , its preference  $\prec_v$  is a strict linear order on  $\delta(v)$ . Let  $N(v)$  be the neighborhood of  $v$ . With a slight abuse of notation, we also view  $\prec_v$  as a strict linear order on  $N(v)$ , such that  $a \prec_v b$  (that is,  $v$  prefers  $a$  to  $b$ ) if and only if  $va \prec_v vb$ . Thus we can naturally associate a *preference list* with  $v$ , which lists its neighbors in the increasing order of  $\prec_v$ . Note that this list is empty if  $N(v) = \emptyset$ . The *preference table* of  $(G, \prec)$  consists of the preference lists of all vertices of  $G$ , and is denoted by  $\mathcal{T}(G, \prec)$  or simply by  $\mathcal{T}(G)$  when  $\prec$  is clear from the context. We use  $f_G(v)$  or  $f_{\mathcal{T}(G)}(v)$  (resp.  $l_G(v)$  or  $l_{\mathcal{T}(G)}(v)$ ) to denote the first (resp. last) vertex on  $v$ 's preference list. A pair of vertices  $\{u, v\}$  is said to *belong to*  $\mathcal{T}(G)$  if  $uv \in E$ . By *deleting* a pair  $\{u, v\}$  from  $\mathcal{T}(G)$  we mean the operation of deleting  $u$  from  $v$ 's preference list and deleting  $v$  from  $u$ 's preference list. For each subgraph  $K$  of  $G$ , let  $\prec^K$  be the projection of  $\prec$  onto  $K$ . Clearly,  $(K, \prec^K)$  is also a preference system. For notational simplicity, we write  $(K, \prec)$  for  $(K, \prec^K)$  hereafter.

In [7], Irving devised a two-phase polynomial-time algorithm for finding a stable matching of  $(G, \prec)$ , if any, or declaring that no such matching exists. It is worthwhile pointing out that Irving's original algorithm was intended for the so-called stable roommates problem, however it can be extended to the general stable matching problem with only a slight modification (see Subsection 4.5.2 of Gusfield and Irving [6]). Now let us present Phase 1 of his algorithm, which is actually a proposal sequence and will play an important role in our proofs.

### Phase 1 of Irving's Algorithm (PHASE I)

**Description:** Initially, the preference table  $\mathcal{T}$  is set to  $\mathcal{T}(G, \prec)$ , and every vertex is set *free*. Each successive proposal in this phase is made by some free vertex  $v$ , with nonempty preference list in  $\mathcal{T}$ , to  $f_{\mathcal{T}}(v)$ , who *receives* the proposal. As a result of this proposal,  $v$  ceases to be free and

becomes *semiengaged* to  $f_{\mathcal{T}}(v)$ , and  $f_{\mathcal{T}}(v)$  *rejects* any vertex who was previously semiengaged to it; the rejected vertex (if any) is reset free. Moreover,  $\mathcal{T}$  is updated by deleting all pairs  $\{f_{\mathcal{T}}(v), u\}$  in  $\mathcal{T}$  with  $v \prec_{f_{\mathcal{T}}(v)} u$ . The phase continues as long as some free vertex has a nonempty preference list (and can therefore make a further proposal).

To facilitate better understanding of this phase, we remark that, first, the semiengagement, as its name implies, is not symmetric. At the point when  $v$  becomes semiengaged to  $f_{\mathcal{T}}(v)$ , this  $f_{\mathcal{T}}(v)$  may be free, or may be semiengaged to someone else; second, there are no “immediate rejections” in the algorithm; such rejections are preempted by the deletion of pairs from the preference table. To be precise, when  $v$  becomes semiengaged to  $f_{\mathcal{T}}(v)$ , all pairs  $\{f_{\mathcal{T}}(v), v'\}$  in that table such that  $f_{\mathcal{T}}(v)$  prefers  $v$  to  $v'$  are deleted, so that  $v$  becomes the last entry on  $f_{\mathcal{T}}(v)$ ’s list, as well as  $f_{\mathcal{T}}(v)$  being the first on  $v$ ’s list. If some other  $u$  was previously semiengaged to  $f_{\mathcal{T}}(v)$ , then it can be shown that this semiengagement is broken, the pair  $\{f_{\mathcal{T}}(v), u\}$  is among those deleted, and  $u$  is set free as a result (see page 167 of Gusfield and Irving [6]).

The preference table produced by PHASEI will be referred to as the *phase-1 table*. Although some nondeterminism is involved in PHASEI, the following statement (Lemma 4.2.1 in Gusfield and Irving [6]) asserts that it is of no consequence.

**Lemma 1.4** *For any given preference system, all possible executions of PHASEI yield the same phase-1 table.* ■

Let  $H$  be a spanning subgraph of  $G$ . We call  $(H, \prec)$  the *representation* of  $(G, \prec)$  if the preference table  $\mathcal{T}(H, \prec)$  is precisely the phase-1 table of  $(G, \prec)$ .

Throughout this paper, let  $\mathbb{P}(G, \prec, w)$  stand for the LP relaxation of the MWSMP

$$\begin{aligned} & \text{Maximize} && w^T x && (1.5) \\ & \text{subject to} && x \in FSM(G, \prec) \end{aligned}$$

and let  $\mathbb{D}(G, \prec, w)$  stand for its dual

$$\begin{aligned} & \text{Minimize} && y(V) - z(E) \\ & \text{subject to} && y(u) + y(v) - z(\psi(uv)) \geq w(uv) && \forall uv \in E, && (1.6) \end{aligned}$$

$$y(v) \geq 0 \quad \forall v \in V, \quad (1.7)$$

$$z(e) \geq 0 \quad \forall e \in E, \quad (1.8)$$

Moreover, we propose to call an integral feasible solution  $(y, z)$  of  $\mathbb{D}(G, \prec, w)$  a *stable cover*<sup>1</sup>, whose cost is defined to be  $y(V) - z(E)$ .

Given  $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$ , such that

- each  $C_i$  is a vertex of  $G$  or an edge of  $G$  or a cycle with cyclic preference in  $(G, \prec)$ ,
- each vertex of  $G$  is contained in some  $C_i$ , and

---

<sup>1</sup>This concept is in some sense very similar to  $w$ -stable set defined by Schrijver [13].

•  $C_1, C_2, \dots, C_k$  are pairwise vertex-disjoint,  
we define  $x_{\mathcal{C}} \in \mathbb{R}^E$  as

$$x_{\mathcal{C}}(e) = \begin{cases} 1 & \text{if } e \in \mathcal{C}, \\ 1/2 & \text{if } e \text{ is contained in some cycle in } \mathcal{C}, \\ 0 & \text{otherwise} \end{cases} \quad (1.9)$$

for all  $e \in E$ , and call  $\mathcal{C}$  a *semistable partition* of  $(G, \prec)$  if  $x_{\mathcal{C}} \in FSM(G, \prec)$ . The partition is so named because, as stated before, Tan [14] characterized all preference systems with (perfect) stable matchings in terms of forbidden structures, the so-called stable partitions with odd parties. (To be specific,  $\mathcal{C}$  is called a *stable partition* of  $(G, \prec)$  if for any  $e \in E$  with  $x_{\mathcal{C}}(e) = 0$ , at least one end  $v$  of  $e$  satisfies  $x_{\mathcal{C}}(\varphi(e) \cap \delta(v)) \geq 1$ , and a member  $C_i$  of  $\mathcal{C}$  is called an *odd party* if  $C_i$  has an odd cardinality.) Obviously, every stable partition is a semistable partition, however the converse need not hold. The other motivation for semistable partitions is given below: by Theorem 3.4 in Abeledo and Rothblum [1], if  $x$  is a half-integral point in  $FSM(G, \prec)$ , then the set of edges in  $E_{1/2}(x)$  forms vertex-disjoint cycles in  $G$ , each having cyclic preferences. Combining this observation with Theorem 1.2(i), we see that every vertex of  $FSM(G, \prec)$  naturally corresponds to a semistable partition of  $(G, \prec)$ . Therefore, to ensure the integrality of  $FSM(G, \prec)$ , we may turn to eliminating certain types of semistable partitions of  $(G, \prec)$ .

Set  $E(\mathcal{C}) = \cup_{i=1}^k E(C_i)$  and

$$E_{\mathcal{C}} = \{uv \in E : \exists e_1, e_2, f_1, f_2 \in E(\mathcal{C}) \text{ such that } e_1 \preceq_u uv \preceq_u e_2 \text{ and } f_1 \preceq_v uv \preceq_v f_2\}. \quad (1.10)$$

Let  $G_{\mathcal{C}}$  be the subgraph of  $G$  induced by all edges in  $E_{\mathcal{C}}$ . Now we are ready to present the main result of this paper, which contains a complete characterization of all preference systems  $(G, \prec)$  with totally dual integral  $\pi(G, \prec)$ .

**Theorem 1.5** *Let  $(G, \prec)$  be a preference system with  $G = (V, E)$ , and let  $(H, \prec)$  be its representation. Then the following statements are equivalent:*

- (i)  $H$  is a bipartite graph;
- (ii)  $G_{\mathcal{C}}$  is bipartite for every semistable partition  $\mathcal{C}$  of  $(G, \prec)$ ;
- (iii)  $SM(G, \prec) = FSM(G, \prec)$ ; and
- (iv)  $\pi(G, \prec)$  is totally dual integral.

Moreover, for any  $G$  as described in (i) and any  $w \in \mathbb{Z}^E$ , a maximum-weight stable matching and a minimum-cost stable cover of  $(G, \prec, w)$  can be found in  $O(m^2 \log m)$  time, where  $m = |E|$ .

Five remarks may help to put our theorem in proper perspective: First, Theorem 1.5 clearly generalizes the above Király-Pap theorem. Second, Theorem 1.2 contains a complete characterization of all preference systems  $(G, \prec)$  whose fractional stable matching polytope is integral; that is,  $SM(G, \prec) = FSM(G, \prec)$ . The equivalence of (ii) and (iii) in our theorem also serves

this purpose, however, the bipartition requirement in our structural description is different from the negation of (ii) in Theorem 1.2. Third, from the definition of TDI system and the aforementioned Edmonds-Giles theorem [2], we see that if  $\pi(G, \prec)$  is totally dual integral, then both  $\mathbb{D}(G, \prec, w)$  and  $\mathbb{P}(G, \prec, w)$  have integral optimal solutions for any  $w \in \mathbb{Z}^E$ . Thereby we get a structural characterization of the following min-max relation on stable matchings.

**Corollary 1.6** *Let  $(G, \prec)$  be a preference system with  $G = (V, E)$ , and let  $(H, \prec)$  be its representation. Then the maximum total weight of a stable matching of  $(G, \prec)$  is equal to the minimum cost of a stable cover of  $(G, \prec)$  for every  $w \in \mathbb{Z}^E$  if and only if  $G_{\mathcal{C}}$  is bipartite for every semistable partition  $\mathcal{C}$  of  $(G, \prec)$  if and only if  $H$  is bipartite.*

We point out that this min-max relation closely resembles König's min-max theorem on bipartite matching and the min-max theorem on weighted bipartite edge covers; see Theorems 16.2 and 19.5 in Schrijver [13]. Fourth, in Tan's characterization [14] of all preference systems with stable matchings, the forbidden structures are stable partitions with odd parties, while the obstructions to our min-max relation are semistable partitions with odd cycles. So these two characterizations are essentially in the same spirit. (Since any integral optimal solution to the MWSMP is a solution to the corresponding SMP, and since the MWSMP is *NP*-hard while the SMP is solvable in polynomial time, presumably more graphical structures should be excluded in our study than in Tan's characterization.) Finally, since  $(H, \prec)$  can be produced by PHASEI in  $O(m)$  time, Theorem 1.5 yields a linear-time algorithm for recognizing the scenarios addressed in (ii)–(iv) by testing if  $H$  is a bipartite graph.

A linear system  $Ax \leq b$  is called *totally dual half-integral* (TDI/2) if the minimum in the LP-duality equation (1.1) has a half-integral optimal solution, for every integral vector  $w$  for which the optimum is finite. It is easy to verify that  $Ax \leq b$  is TDI/2 if and only if  $Bx \leq b$  is TDI, where  $B = A/2$ . Thus, from the above Edmonds-Giles theorem [2], we deduce that if  $Ax \leq b$  is TDI/2 and  $b$  is integral, then the maximum in equation (1.1) also has a half-integral optimal solution, for every integral vector  $w$  for which the optimum is finite. Recall Theorem 1.2, for any preference system, all vertices of the fractional stable matching polytope are half-integral. Clearly, our next theorem strengthens this result.

**Theorem 1.7** *Let  $(G, \prec)$  be a preference system with  $G = (V, E)$ . Then  $\pi(G, \prec)$  is totally dual half-integral. Moreover, for any  $w \in \mathbb{Z}^E$ , a half-integral optimal solution to  $\mathbb{P}(G, \prec, w)$  and a half-integral optimal solution to  $\mathbb{D}(G, \prec, w)$  can be found in  $O(m^2 \log m)$  time, where  $m = |E|$ .*

## 2 Preliminaries

In our proofs, we shall use the following charming property enjoyed by stable matchings, which was first exhibited by Abeledo and Rothblum [1] (see Theorem 4.2).

**Theorem 2.1** (Abeledo-Rothblum [1]) *For any preference system  $(G, \prec)$ , the vertex set of  $G$  can be partitioned into  $V^0$  and  $V^1$  such that for every  $x \in FSM(G, \prec)$ , there holds  $x(\delta(v)) = i$  for all  $v \in V^i$  and  $i = 0, 1$ . (So all vertices in  $V^1$  are matched in every stable matching of  $(G, \prec)$ , and no vertex in  $V^0$  is matched in any stable matching.)  $\blacksquare$*

Now let us exhibit some properties enjoyed by the representation of a preference system. As usual, we use  $V(K)$  and  $E(K)$  to denote the vertex set and edge set of a graph  $K$ , respectively.

**Lemma 2.2** *Let  $(G, \prec)$  be a preference system with  $G = (V, E)$ , let  $(H, \prec)$  be its representation, and let  $V^1$  be the set of all non-isolated vertices of  $H$ . Then the following statements hold:*

- (i)  *$(G, \prec)$  and  $(H, \prec)$  have the same set of stable matchings. Besides,  $V^1$  consists of exactly the matched vertices in every stable matching (if any) of  $(G, \prec)$  and  $(H, \prec)$ ;*
- (ii) *For any distinct  $u, v \in V$ ,  $u = f_H(v)$  if and only if  $v = l_H(u)$ ;*
- (iii) *An edge  $uv \in E - E(H)$  if and only if  $u \in V^1$  prefers  $l_H(u)$  to  $v$  or  $v \in V^1$  prefers  $l_H(v)$  to  $u$ ;*
- (iv) *For each  $v \in V^1$ , if  $f_H(v) \neq l_H(v)$ , then  $f_H(f_H(v)) \neq v$ ; and*
- (v) *For each  $v \in V^1$ , let  $G_v$  be the graph obtained from  $G$  by deleting the edge  $vl_G(v)$ . If  $H \subseteq G_v$ , then  $(H, \prec)$  is also the representation of  $(G_v, \prec)$ .*

**Proof.** The first three statements can all be found in Gusfield and Irving [6]: the first half of (i) is exactly the same as Lemma 4.2.3(i), the second half of (i) is contained in Theorem 4.5.2, and statements (ii) and (iii) are the subjects of Lemma 4.2.2.

To justify (iv), observe from PHASEI that every vertex  $u \in V^1$  has to propose to  $f_H(u)$  at some step and remains semiengaged to  $f_H(u)$  from that step onwards, otherwise the pair  $\{u, f_H(u)\}$  would get deleted in the algorithm at some other step and hence  $uf_H(u)$  would not be an edge of  $H$ , a contradiction. Assume on the contrary that  $f_H(f_H(v)) = v$  while  $f_H(v) \neq l_H(v)$  for some  $v \in V^1$ . Then right after  $f_H(v)$  proposed to  $v$ , the pair  $\{v, l_H(v)\}$  would get deleted in PHASEI because by hypothesis  $f_H(v) \prec_v l_H(v)$ , contradicting the fact that  $vl_H(v) \in E(H)$ . So (iv) is established.

To prove (v), note that since  $vl_G(v)$  is not an edge of  $H$  and since  $v \in V^1$ , we have  $vf_H(v) \prec_v vl_G(v)$ . Thus  $v$  has never proposed to  $l_G(v)$  in PHASEI.

CLAIM: There exists an execution of PHASEI in which  $l_G(v)$  has never proposed to  $v$ .

Assume the contrary:  $l_G(v)$  has to propose to  $v$  in all possible executions of PHASEI. Let us consider an execution, denoted by  $(\alpha)$ , where the step at which  $l_G(v)$  proposes to  $v$  is as late as possible. Observe that after  $l_G(v)$  proposed to  $v$ , nothing was done except for the semiengagement of  $l_G(v)$  to  $v$ . Moreover, no more proposal was possible (for otherwise, let  $k$  be the index such that the  $k$ th proposal in  $(\alpha)$  is the one made by  $l_G(v)$  to  $v$ , and let  $(\beta)$  be an execution whose first  $k - 1$  proposals are exactly the same as those of  $(\alpha)$ , while  $k$ th proposal is the  $(k + 1)$ st proposal of  $(\alpha)$ . Then we can see that either  $l_G(v)$  has never proposed to  $v$  in  $(\beta)$  or



the step at which  $l_G(v)$  proposes to  $v$  in  $(\beta)$  is later than that in  $(\alpha)$ , so we reach a contradiction in either case). Thus execution  $(\alpha)$  terminates right after the proposal made by  $l_G(v)$  to  $v$ . It follows that  $(\alpha)$  yields a phase-1 table which contains  $\{v, l_G(v)\}$  as a pair, contradicting Lemma 1.4 because  $vl_G(v)$  is not an edge of  $H$ . Thus the claim is justified.

Let  $(\gamma)$  be an execution of PHASEI as claimed above. Since  $v$  and  $l_G(v)$  have never proposed to each other in  $(\gamma)$ , and since  $(\gamma)$  terminates with  $\mathcal{T}(H, \prec)$  as the phase-1 table,  $(\gamma)$  naturally corresponds to an execution of PHASEI on  $G_v$  which terminates with  $\mathcal{T}(H, \prec)$  as the phase-1 table as well. It follows that  $(H, \prec)$  is also the representation of  $(G_v, \prec)$ . Hence (v) holds. ■

In view of Lemma 2.2(ii), we introduce the following notation:

$$F(H) = \{vf_H(v) : v \in V^1\} = \{vl_H(v) : v \in V^1\}. \quad (2.1)$$

Let  $\hat{H}$  be the subgraph of  $H$  induced by all edges in  $F(H)$ . Thus  $\hat{H} = (V^1, F(H))$ . Throughout we use  $\mathfrak{S}$  to denote the collection of all preference systems  $(G, \prec)$  such that  $G_{\mathcal{C}}$  is bipartite for every semistable partition  $\mathcal{C}$  of  $(G, \prec)$  (recall the definition above Theorem 1.5).

**Lemma 2.3** *Let  $G, H$ , and  $V^1$  be as given in Lemma 2.2, let  $F(H)$  and  $\hat{H}$  be as defined above, and let  $V^0 = V - V^1$ . Then the following statements hold:*

- (i)  $\hat{H}$  is the disjoint union of some edges and cycles, such that each of these cycles has a cyclic preference in  $(G, \prec)$ ;
- (ii) Let  $\mathcal{D}$  be the set of all edges and cycles as specified in (i). Then each edge in  $E(H) - F(H)$  (if any) has both ends on cycles in  $\mathcal{D}$ ;
- (iii) Let  $\mathcal{C}$  be obtained from  $\mathcal{D}$  by adding all vertices in  $V^0$  as elements. Then  $\mathcal{C}$  is a semistable partition of  $(G, \prec)$ ; and
- (iv)  $H \setminus V^0 = G_{\mathcal{C}}$ , which implies that  $H$  is a bipartite graph if  $(G, \prec) \in \mathfrak{S}$ .

**Proof.** To justify (i), note that, by Lemma 2.2(ii), the maximum degree of  $\hat{H}$  is at most two, and every cycle in  $\hat{H}$  has a cyclic preference in  $(G, \prec)$ . If the assertion fails, then  $\hat{H}$  would contain a longest induced path  $P = v_1v_2v_3\dots v_k$ , with  $k \geq 3$ . Lemma 2.2(ii) allows us to assume that  $f_H(v_i) = v_{i-1}$  and  $l_H(v_i) = v_{i+1}$  for  $i = 2, 3, \dots, k-1$ . Thus  $f_H(v_2) = v_1 \neq v_3 = l_H(v_2)$ . By Lemma 2.2(iv) (with  $v_2$  in place of  $v$ ), we have  $v_2 \neq f_H(f_H(v_2)) = f_H(v_1)$ . Since  $v_1f_H(v_1)$  is contained in  $F(H)$ , it is an edge of  $\hat{H}$ . Hence the degree of  $v_1$  in  $\hat{H}$  is precisely two, contradicting the assumption that  $P$  is a longest induced path in  $\hat{H}$ . So (i) holds.

To justify (ii), let  $uv \in E(H) - F(H)$ . From (2.1), we deduce that  $uf_H(u) \prec_u uv \prec_u ul_H(u)$  and  $vf_H(v) \prec_v uv \prec_v vl_H(v)$ . Thus  $f_H(u) \neq l_H(u)$  and  $f_H(v) \neq l_H(v)$ . By definition, both  $u$  and  $v$  have degree two in  $\hat{H}$ . Hence (ii) follows instantly from (i).

To justify (iii), recall (1.9) and (i). Clearly  $x = x_{\mathcal{C}}$  satisfies (1.2) and (1.4). It remains to show that  $x$  satisfies (1.3); that is,  $x(\varphi(e)) \geq 1$  for any  $e \in E$ . To this end, let  $e = uv$ . For  $e \in F(H)$ , by (2.1), we may assume that  $v \in V^1$  and  $u = l_H(v)$ . Since  $vf_H(v) \preceq_v vl_H(v) = e$ , we

have  $x(\varphi(e)) \geq x(e) = 1$  if  $e \in \mathcal{C}$  and  $x(\varphi(e)) \geq x(e) + x(vf_H(v)) = 1/2 + 1/2 = 1$  otherwise. For  $e \in E(H) - F(H)$ , it follows from (ii) that  $x(\varphi(e)) \geq x(uf_H(u)) + x(vf_H(v)) = 1/2 + 1/2 = 1$ . For  $e \in E - E(H)$ , by Lemma 2.2(iii), there exists  $t \in V^1 \cap \{u, v\}$  such that  $tl_H(t) \prec_t e$ , yielding  $x(\varphi(e)) \geq x(\delta(t) \cap F(H)) = 1$ . So  $x$  satisfies (1.3) as well. By definition,  $\mathcal{C}$  is a semistable partition of  $(G, \prec)$ . This proves (iii).

To justify (iv), we appeal to Lemma 2.2(iii): an edge  $uv$  of  $G$  is contained in  $H$  if and only if  $uf_H(u) \preceq_u uv \preceq_u ul_H(u)$  and  $vf_H(v) \preceq_v uv \preceq_v vl_H(v)$ . In view of (i), (iii) and (1.10), we obtain  $E(H) = E_{\mathcal{C}}$ . So (iv) is established.  $\blacksquare$

Our next lemma implies that  $V^i$ , for  $i = 0, 1$ , involved in Lemma 2.3 is precisely the same as that in Theorem 2.1.

**Lemma 2.4** *Let  $G, H, V^0$ , and  $V^1$  be as specified in the above two lemmas. Then for every  $x \in FSM(G, \prec)$ , the following statements hold:*

- (i)  $x(\delta(v)) = i$  for every  $v \in V^i$  and  $i = 0, 1$ ; and
- (ii)  $x(E - E(H)) = 0$ , which implies that  $x|_{E(H)} \in FSM(H, \prec)$ .

**Proof.** Let  $\mathcal{C}$  be the semistable partition of  $(G, \prec)$  as specified in Lemma 2.3(iii). Since  $x_{\mathcal{C}} \in FSM(G, \prec)$ , from (1.9) we deduce that  $x_{\mathcal{C}}(\delta(v)) = i$  for  $i = 0, 1$  if and only if  $v \in V^i$ . Thus (i) follows from Theorem 2.1.

Assume on the contrary that  $x(uv) > 0$  for some  $x \in FSM(G, \prec)$  and  $uv \in E - E(H)$ ; subject to this, we further assume that no pair  $\{r, s\}$  with  $x(rs) > 0$  had been deleted in PHASE I before  $\{u, v\}$ . Renaming  $u$  and  $v$  if necessary, we may suppose  $\{u, v\}$  was deleted from the preference table  $\mathcal{T}$  because  $u$  received a proposal from vertex  $t$ . From PHASE I, we see that  $ut \prec_u uv$ . In view of the choice of  $uv$ , no edge in  $\delta(t) \cap \{e \in E : x(e) > 0\}$  had been deleted at the step when  $t$  proposed to  $u$ . So  $ut$  dominates all edges in  $\delta(t) \cap \{e \in E : x(e) > 0\}$  in  $(G, \prec)$ . It follows that  $x(\varphi(ut)) \leq x(\delta(u)) - x(uv)$ . By (1.2), we have  $x(\delta(u)) \leq 1$ , which implies  $x(\varphi(ut)) \leq 1 - x(uv) < 1$ , contradicting (1.3) for  $x \in FSM(G, \prec)$ . Thus (ii) also holds.  $\blacksquare$

We shall establish the main result of this paper by an induction method; the following statement will be used to dictate the induction step.

**Lemma 2.5** *Let  $(G, \prec)$  be a preference system with  $G = (V, E)$ , and let  $(H, \prec)$  be its representation with  $E - E(H) \neq \emptyset$ . Then  $G$  has three distinct vertices  $r, s, t$ , such that  $f_G(r) = s \in V^1$ ,  $l_G(s) = t$ , and  $st \in E - E(H)$ .*

**Proof.** Set  $F = \{vf_G(v) : v \in V \text{ and } N(v) \neq \emptyset\}$ . If  $F \subseteq E(H)$ , then PHASE I consists of  $|F|$  proposals resulting in no rejections, and terminates with  $F(H) = F$ . By Lemma 2.2(iii), there exists an edge  $su \in E - E(H)$ , such that  $s \in V^1$  and  $sl_H(s) \prec_s su$ . Set  $r = l_H(s)$  and  $t = l_G(s)$ . Then  $sr \prec_s su \preceq_s st$ . From the definition of  $l_H(s)$ , we see that  $st \in E - E(H)$ . By Lemma

2.2(ii),  $s = f_H(r)$ . As  $F \subseteq E(H)$ , we have  $s = f_G(r)$ . Thus  $r, s, t$  are as desired, and hence we may assume that  $F - E(H) \neq \emptyset$ .

Set  $J = \{v \in V : vf_G(v) \in F - E(H)\}$ . Let us consider an execution of PHASEI in which every free vertex  $v$  in  $J$  has the *priority* to propose to  $f_G(v)$  who is still on  $v$ 's preference list (breaking ties arbitrarily). Then, under the assumption  $F - E(H) \neq \emptyset$ , some vertex in  $J$  would certainly be rejected by its most preferred neighbor in  $G$  at some step. Suppose the earliest rejection happened when a vertex  $v \in J$  was rejected by  $s = f_G(v)$  upon receiving proposal from some  $r \in V - \{v\}$ . Since no vertex in  $J$  had been rejected before, no pair which corresponds to an edge in  $F$  had been deleted, and hence the proposal made by  $r$  was to  $s = f_G(r)$ , which implies that  $f_G(r) = f_G(v)$  and  $\{rs, vs\} \subseteq \delta(s) \cap F$ . Since  $s$  rejected  $v$  due to the proposal made by  $r$ , we obtain  $sr \prec_s sv \preceq_s sl_G(s)$ . From the description of PHASEI, it can be seen that if a vertex receives a proposal in an execution, then it belongs to  $V^1$ . Hence  $s \in V^1$ . Set  $t = l_G(s)$ . Clearly  $r, s, t$  are as desired.  $\blacksquare$

### 3 Proofs of Theorems

We break the proof of Theorem 1.5 into a series of lemmas.

**Lemma 3.1** *Let  $(G, \prec)$  be a preference system. If  $FSM(G, \prec) = SM(G, \prec)$ , then  $(G, \prec) \in \mathfrak{S}$ .*

**Proof.** Let  $\mathcal{C}$  be a semistable partition of  $(G, \prec)$ . We aim to prove that  $G_{\mathcal{C}}$  is a bipartite graph. Since  $x_{\mathcal{C}} \in FSM(G, \prec)$  (see (1.9)), by hypothesis we further have  $x_{\mathcal{C}} \in SM(G, \prec)$ . Therefore there exist stable matchings  $M^{(i)}$  of  $(G, \prec)$ ,  $i = 1, 2, \dots, k$ , such that  $x_{\mathcal{C}}$  is a convex combination of the incidence vectors  $x^{(i)}$  of  $M^{(i)}$ . It follows that  $\cup_{i=1}^k M^{(i)} = E(\mathcal{C})$ . In view of Theorem 2.1, the vertex subsets of  $G$  matched by all  $M^{(i)}$  are the same, so each  $M^{(i)}$  is a perfect matching of  $K$ , the subgraph of  $G$  induced by all edges in  $E(\mathcal{C})$ , which in turn implies that all cycles in  $\mathcal{C}$  are even. From (1.10), it can be seen that

$$E(K) = E(\mathcal{C}) = \{vf_{G_{\mathcal{C}}}(v) : v \in V(K)\} = \{vl_{G_{\mathcal{C}}}(v) : v \in V(K)\}. \quad (3.1)$$

Let  $W_1$  be the independent set in  $K$  such that  $M^{(1)} = \{vf_{G_{\mathcal{C}}}(v) : v \in W_1\}$  and let  $W_2 = V(K) - W_1$ . As  $x_{\mathcal{C}} \in FSM(G, \prec)$ , we deduce that  $2x_{\mathcal{C}} - x^{(1)}$  is the incidence vector of the stable matching  $\{vf_{G_{\mathcal{C}}}(v) : v \in W_2\}$  of  $(G, \prec)$ . So we may assume that  $M^{(2)} = \{vf_{G_{\mathcal{C}}}(v) : v \in W_2\}$  and  $x_{\mathcal{C}} = \frac{1}{2}(x^{(1)} + x^{(2)})$ . If some edge  $uv \in E_{\mathcal{C}}$  has both ends in  $W_i$  for  $i = 1$  or  $2$ , then by (3.1) the edges in  $M^{(3-i)}$  incident with  $u$  and  $v$  are  $ul_{G_{\mathcal{C}}}(u)$  and  $vl_{G_{\mathcal{C}}}(v)$ , respectively, neither dominating  $uv$ , which contradicts the stability of  $M^{(3-i)}$ . Thus each edge in  $G_{\mathcal{C}}$  has one end in  $W_1$  and the other in  $W_2$ , and hence  $G_{\mathcal{C}}$  is a bipartite graph, as desired.  $\blacksquare$

**Lemma 3.2** *Let  $(G, \prec)$  be a preference system with a bipartite representation  $(H, \prec)$  (that is,  $H$  is bipartite). Then  $\pi(G, \prec)$  is totally dual integral.*

We present two proofs of this lemma; the first one was discovered by Tamás Király, which is shorter but does not yield a polynomial-time algorithm for the minimum-cost stable cover problem, while the second is the backbone of subsequent Algorithm 3.5.

Király's proof is based on the following technical lemma.

**Lemma 3.3** *Let*

$$Ax + B\bar{x} \leq b, x \geq \mathbf{0}, \bar{x} \geq \mathbf{0} \quad (3.2)$$

and

$$Ax \leq b, x \geq \mathbf{0} \quad (3.3)$$

be two linear systems. Suppose  $\bar{x} \equiv \mathbf{0}$  for all feasible solutions of (3.2). If (3.3) is totally dual integral, then so is (3.2).

**Proof.** Let  $w$  (resp.  $\bar{w}$ ) be an arbitrary integral vector with the same length as  $x$  (resp.  $\bar{x}$ ). By hypothesis, (3.3) is a TDI system, so the linear program  $\min\{y^T b : y^T A \geq w^T, y \geq \mathbf{0}\}$  has an integral optimal solution  $y^*$ . Let  $c$  be the vector obtained from  $\bar{w} - B^T y^*$  by replacing each negative entry with 0. Then  $c \geq \bar{w} - B^T y^*$ . As the optimal value of the linear program

$$\max\{\mathbf{0}^T x + c^T \bar{x} : x \text{ and } \bar{x} \text{ satisfy (3.2)}\}$$

is zero, its dual has an optimal solution  $\bar{y}$  satisfying  $\bar{y}^T A \geq \mathbf{0}$ ,  $\bar{y}^T B \geq c^T$ ,  $\bar{y} \geq \mathbf{0}$ , and  $\bar{y}^T b = 0$ . This zero optimal value and the fact  $c \geq \mathbf{0}$  allow us to assume that  $\bar{y}$  is integral (otherwise replace  $\bar{y}$  with  $k\bar{y}$  for some positive integer  $k$ ). Set  $\bar{y}^* = y^* + \bar{y}$ . Clearly,  $\bar{y}^*$  is integral and satisfies

$$(\bar{y}^*)^T A \geq (y^*)^T A \geq w^T, (\bar{y}^*)^T B \geq (y^*)^T B + c^T \geq \bar{w}^T, \bar{y}^* \geq \mathbf{0}, \text{ and } (\bar{y}^*)^T b = (y^*)^T b. \quad (3.4)$$

Since  $\bar{x} \equiv \mathbf{0}$  for all feasible solutions of (3.2), and  $y^*$  is the optimal solution to the dual of  $\max\{w^T x : x \text{ satisfies (3.3)}\}$ , we deduce that

$$\max\{w^T x + \bar{w}^T \bar{x} : x \text{ and } \bar{x} \text{ satisfy (3.2)}\} = \max\{w^T x : x \text{ satisfies (3.3)}\} = (y^*)^T b,$$

which, together with (3.4), implies that  $\bar{y}^*$  is an integral optimal solution to the dual of  $\max\{w^T x + \bar{w}^T \bar{x} : x \text{ and } \bar{x} \text{ satisfy (3.2)}\}$ , and hence (3.2) is a TDI system.  $\blacksquare$

**Király's Proof of Lemma 3.2.** Let  $G = (V, E)$ . Write  $\pi(G, \prec)$  as  $Ax|_{E(H)} + Bx|_{E-E(H)} \leq b$ ,  $x|_{E(H)} \geq \mathbf{0}$ ,  $x|_{E-E(H)} \geq \mathbf{0}$ , where the columns of  $A$  and  $B$  are indexed by edges in  $E(H)$  and in  $E - E(H)$ , respectively, and  $b|_V = \mathbf{1}$ ,  $b|_E = -\mathbf{1}$ . By Lemma 2.4(ii), we have

$$\begin{aligned} x|_{E-E(H)} &\equiv \mathbf{0} \text{ for every feasible solution } x \in \mathbb{R}^E \text{ to} \\ Ax|_{E(H)} + Bx|_{E-E(H)} &\leq b, x|_{E(H)} \geq \mathbf{0}, x|_{E-E(H)} \geq \mathbf{0}. \end{aligned} \quad (3.5)$$

Let  $w$  be an arbitrary vector in  $\mathbb{Z}^{E(H)}$ , let  $L = V \cup E(H)$ , and let  $C$  be the submatrix of  $A$  formed by rows corresponding to elements in  $L$ . With a slight abuse of the notation, put  $x = x|_{E(H)}$ . Applying Theorem 1.3 to the bipartite preference system  $(H, \prec)$ , we see that  $Cx \leq b|_L, x \geq \mathbf{0}$  is a TDI system. So  $\max\{w^T x : Cx \leq b|_L, x \geq \mathbf{0}\}$  and its dual have integral optimal solutions  $x^*$  and  $y^*$ , respectively. By Lemma 2.2(i),  $x^*$  satisfies  $Ax \leq b, x \geq \mathbf{0}$ , which implies

$$\max\{w^T x : Ax \leq b, x \geq \mathbf{0}\} = \max\{w^T x : Cx \leq b|_L, x \geq \mathbf{0}\}. \quad (3.6)$$

Extending  $y^* \in \mathbb{Z}_+^L$  to  $y \in \mathbb{Z}_+^{V \cup E}$  by adding zero entries, we obtain  $y^T A = (y^*)^T C \geq w^T, y \geq \mathbf{0}$ , and  $y^T b = (y^*)^T b|_L = \max\{w^T x : Cx \leq b|_L, x \geq \mathbf{0}\}$ . From (3.6), we conclude that  $y$  is an integral optimal solution to the dual of  $\max\{w^T x : Ax \leq b, x \geq \mathbf{0}\}$ . Hence  $Ax \leq b, x \geq \mathbf{0}$  is a TDI system. From (3.5) and Lemma 3.3, we thus deduce that  $\pi(G, \prec)$  is also totally dual integral.  $\blacksquare$

**Second Proof of Lemma 3.2.** Let  $G = (V, E)$ . To establish the assertion, we need to show that

(1)  $\mathbb{D}(G, \prec, w)$  has an integral optimal solution for all  $w \in \mathbb{Z}^E$ .

To this end, we apply induction on  $|E|$ . If  $|E| = 0$ , then  $E = \emptyset$  and thus (1) is trivial. So we proceed to the induction step, and assume that (1) holds for all preference systems with fewer edges in the corresponding graphs.

If  $G = H$ , then (1) follows instantly from Theorem 1.3. So we assume that  $G \neq H$  and hence  $E - E(H) \neq \emptyset$ . Let  $V^1$  be the set of all non-isolated vertices of  $H$ . Using Lemma 2.5, we obtain

(2)  $G$  has three distinct vertices  $r, s, t$ , such that  $f_G(r) = s \in V^1, l_G(s) = t$ , and  $st \in E - E(H)$ .

From (2), it can be seen that

(3)  $e \preceq_s st$  for all  $e \in \delta(s)$  and hence  $\psi(st) = \{st\} \cup \{e \in \delta(t) : st \prec_t e\}$ .

Let  $(G', \prec)$  be the preference system whose preference table is obtained from  $\mathcal{T}(G, \prec)$  by deleting the pair  $\{s, t\}$ , with  $G' = (V, E')$ . Observe that  $G' = G \setminus st$  and  $E' = E - \{st\}$ . From (2) and Lemma 2.2(v), we see that

(4)  $(H, \prec)$  is also the representation of  $(G', \prec)$ .

In the remainder of our proof, we use  $\delta'(v)$  to denote the set of all edges of  $G'$  incident with a vertex  $v$ , and use  $\varphi'(e)$  (resp.  $\psi'(e)$ ) to denote the set of all edges of  $G'$  dominating (resp. dominated by) an edge  $e$  in  $E'$ . In view of Lemma 2.3(iii) (with  $(G', \prec)$  in place of  $(G, \prec)$ ), we have  $FSM(G', \prec) \neq \emptyset$  (actually this result was first established for all preference systems by Abeledo and Rothblum [1] using a different method). Motivated by Lemma 2.4(ii), we propose to show that

(5) Let  $\bar{x}$  be a vector in  $FSM(G', \prec)$  and let  $x$  be obtained from  $\bar{x}$  by adding one more entry  $x(st) = 0$ . Then  $x \in FSM(G, \prec)$ .

To justify this, note that

- $x(\delta(v)) = \bar{x}(\delta'(v)) \leq 1$  for all  $v \in V$ , and
- $x(\varphi(e)) = \bar{x}(\varphi'(e)) \geq 1$  for all  $e \in E'$ .

For the edge  $st$ , by (3) we have  $e \prec_s st$  for all  $e \in \delta'(s)$ , which implies that  $x(\varphi(st)) \geq \bar{x}(\delta'(s))$ . Since  $s \in V^1$ , by (4) and Lemma 2.4(i) (with  $G'$  in place of  $G$ ), we obtain  $\bar{x}(\delta'(s)) = 1$ . So  $x(\varphi(st)) \geq 1$ . Combining the above observations, we conclude that  $x$  satisfies (1.2)–(1.4) simultaneously, so (5) holds.

By (4) and induction hypothesis on  $(G', \prec)$ , the linear system  $\pi(G', \prec)$  is totally dual integral. Thus the definition of TDI system guarantees the existence of an integral optimal solution  $(\bar{y}, \bar{z})$  to  $\mathbb{D}(G', \prec, \bar{w})$ , where  $\bar{w} = w|_{E'}$ . Let  $\bar{x}$  be an optimal solution to  $\mathbb{P}(G', \prec, \bar{w})$ . It follows from the LP duality theorem that

$$(6) \quad \bar{y}(V) - \bar{z}(E') = \bar{w}^T \bar{x}.$$

Let  $x^*$  be the vector obtained from  $\bar{x}$  by adding one more entry  $x^*(st) = 0$ . From (5), we see that

$$(7) \quad x^* \text{ is a feasible solution to } \mathbb{P}(G, \prec, w).$$

Set  $\epsilon = \max\{0, w(st) - \bar{y}(s) - \bar{y}(t) + \bar{z}(\psi(st) - \{st\})\}$ . Let  $y^*$  be the vector obtained from  $\bar{y}$  by replacing  $\bar{y}(s)$  with  $\bar{y}(s) + \epsilon$ , and let  $z^*$  be the vector obtained from  $\bar{z}$  by replacing  $\bar{z}(rs)$  with  $\bar{z}(rs) + \epsilon$  and then adding one more entry  $z^*(st) = 0$ . By (6), we have  $y^*(V) - z^*(E) = \bar{y}(V) + \epsilon - (\bar{z}(E') + \epsilon) = \bar{w}^T \bar{x}$ . Hence

$$(8) \quad y^*(V) - z^*(E) = w^T x^*.$$

We propose to show that

$$(9) \quad y^*(u) + y^*(v) - z^*(\psi(uv)) \geq w(uv) \text{ for all } uv \in E.$$

To justify this, we distinguish among five cases.

**Case 1.**  $\{u, v\} \cap \{r, s, t\} = \emptyset$ . In this case,  $y^*(u) + y^*(v) - z^*(\psi(uv)) = \bar{y}(u) + \bar{y}(v) - \bar{z}(\psi'(uv)) \geq w(uv)$  because  $(\bar{y}, \bar{z})$  is a feasible solution to  $\mathbb{D}(G', \prec, \bar{w})$ .

**Case 2.**  $u = s$  and  $v = t$ . In this case,  $y^*(s) + y^*(t) - z^*(\psi(st)) = (\bar{y}(s) + \epsilon) + \bar{y}(t) - z^*(st) - z^*(\psi(st) - \{st\}) = \bar{y}(s) + \epsilon + \bar{y}(t) - \bar{z}(\psi(st) - \{st\}) \geq w(st)$  by the definition of  $\epsilon$ .

**Case 3.**  $u = s$  and  $v \neq t$ . In this case,  $z^*(\psi(sv)) = \bar{z}(\psi'(sv) - \{sr\}) + z^*(sr) \cdot |\{sr\} \cap \psi'(sv)| + z^*(st) = \bar{z}(\psi'(sv) - \{sr\}) + (\bar{z}(sr) + \epsilon) \cdot |\{sr\} \cap \psi'(sv)| \leq \bar{z}(\psi'(sv)) + \epsilon$ . So  $y^*(s) + y^*(v) - z^*(\psi(sv)) \geq (\bar{y}(s) + \epsilon) + \bar{y}(v) - (\bar{z}(\psi'(sv)) + \epsilon) = \bar{y}(s) + \bar{y}(v) - \bar{z}(\psi'(sv)) \geq w(sv)$ .

**Case 4.**  $u \neq s$  and  $v = t$ . In this case,  $\psi(ut) \subseteq \psi'(ut) \cup \{st\}$  and  $y^*(u) + y^*(t) - z^*(\psi(ut)) \geq \bar{y}(u) + \bar{y}(t) - (\bar{z}(\psi'(ut)) + z^*(st)) = \bar{y}(u) + \bar{y}(t) - \bar{z}(\psi'(ut)) \geq w(ut)$ .

**Case 5.**  $u = r$  and  $v \neq s$ . In this case, we have  $rs \notin \psi(rv) = \psi'(rv)$  for  $s = f_G(r)$ , which implies that  $z^*(\psi(rv)) = \bar{z}(\psi'(rv))$ . So  $y^*(r) + y^*(v) - z^*(\psi(rv)) = \bar{y}(r) + \bar{y}(v) - \bar{z}(\psi'(rv)) \geq w(rv)$ .

As these cases have exhausted all possibilities, (9) is established.

From (9), it follows that

$$(10) \quad (y^*, z^*) \text{ is an integral feasible solution to } \mathbb{D}(G, \prec, w).$$

Combining (7), (8) and (10) and using the LP duality theorem, we can further conclude that  $(y^*, z^*)$  is an integral optimal solution to  $\mathbb{D}(G, \prec, w)$ . This proves (1) and hence Lemma 3.2. ■

As described in Gusfield and Irving [6] and Király and Pap [8], network flow techniques can be used to solve various stable matching problems and their duals. The following lemma is concerned with their algorithm for finding a maximum-weight stable matching and a minimum-cost stable cover of a bipartite preference system. Since the algorithm is lengthy, we shall neither get into the details nor discuss its correctness. What we are going to do is to outline its major steps and figure out its complexity, which is not given explicitly in [6] or [8].

**Lemma 3.4** [6, 8] *Let  $(G, \prec)$  be a bipartite preference system with  $G = (V, E)$ . Then for any  $w \in \mathbb{Z}^E$ , a maximum-weight stable matching and a minimum-cost stable cover of  $(G, \prec, w)$  can be found in  $O(m^2 \log m)$  time, where  $m = |E|$ .*

**Proof.** To establish this complexity result, let us give a sketch of their algorithm. In the initialization step, we construct the following objects (see [8] for undefined terms):

- (i) the set  $R$  of all rotations of  $(G, \prec)$ , such that each  $\rho \in R$  is expressed in the form  $\rho = (v_1, u_1, v_2, u_2, \dots, v_k, u_k)$  for which  $w^*(\rho)$  is defined to be  $\sum_{i=1}^k (-w(v_i u_i) + w(u_i v_{i+1}))$ , with  $v_{k+1} = v_1$ ;
- (ii) the rotation digraph  $D = (R, A)$  whose closed subsets (i.e. vertex subsets with no incoming arcs from the outside) correspond to stable matchings of  $(G, \prec)$ ; and
- (iii) the so-called man-optimal stable matching  $M_0$  of  $(G, \prec)$ .

Since both  $|R|$  and  $|A|$  are linear in  $m$  (see [8] and proof of Lemma 3.3.2 in [6]), as discussed in Subsection 3.3.1 and Section 3.6 of [6], this step can be carried out in  $O(nm)$  time, where  $n = |V|$ .

Let  $opt$  denote the optimal objective value of the following problem

$$\begin{aligned} & \text{Maximize} && (w^*)^T \beta \\ & \text{subject to} && \beta(\rho) - \beta(\rho') \geq 0 && \forall (\rho, \rho') \in A, \\ & && 1 \geq \beta(\rho) \geq 0 && \forall \rho \in R \end{aligned}$$

(see (i) for the definition of  $w^*$ ) and its dual

$$\begin{aligned} & \text{Minimize} && \gamma(R) \\ & \text{subject to} && \gamma(\rho) - \gamma(A^+(\rho)) + \gamma(A^-(\rho)) \geq w^*(\rho) && \forall \rho \in R, \gamma \in \mathbb{R}_+^{R \cup A}, \end{aligned} \quad (3.7)$$

where  $A^+(\rho)$  and  $A^-(\rho)$  are the sets of arcs in  $D$  leaving  $\rho$  and entering  $\rho$ , respectively. In the main step of the algorithm, our objective is to find a closed subset  $S$  of  $R$  with  $w^*(S) = opt$  and an integral optimal solution  $\gamma^*$  to problem (3.7). As elaborated in the proofs of Theorem 3.6.2

and Lemma 3.6.3 in [6], the desired  $S$  can be obtained from a minimum cut in an auxiliary flow network  $\vec{D}$  (constructed from  $D$  by adding a source and a sink). It is not difficult to see that an integral maximum flow in  $\vec{D}$  also yields the desired  $\gamma^*$ . So this step can be carried out in  $O(m^2 \log m)$  time using the Goldberg-Tarjan algorithm [5, 13].

Given such an  $S$ , we can finally obtain a stable matching of  $(G, \prec, w)$  with maximum possible weight  $w(M_0) + \text{opt}$  from  $M_0$  by eliminating all rotations in  $S$  in  $O(nm)$  time (see [8] or Subsections 3.6.1 and 3.6.2 of [6] for details). Besides, starting from  $\gamma^*$ , as shown by Király and Pap [8], it takes  $O(nm)$  time to find a minimum-cost stable cover of  $(G, \prec, w)$ .

Therefore, a maximum-weight stable matching and a minimum-cost stable cover of  $(G, \prec, w)$  can be found in  $O(m^2 \log m)$  time.  $\blacksquare$

Now let us present an efficient combinatorial algorithm for finding a maximum-weight stable matching and a minimum-cost stable cover in a more general setting, where  $\delta_i(t)$  is the set of all edges in  $G_i$  incident with the vertex  $t$ .

---

**Algorithm 3.5** FOR THE MAXIMUM-WEIGHT STABLE MATCHING PROBLEM & ITS DUAL

---

**Input:** A preference system  $(G, \prec)$  with a bipartite representation  $(H, \prec)$  and a weight function  $w \in \mathbb{Z}^E$ .

**Output:** A maximum-weight stable matching  $M$  and a minimum-cost stable cover  $(y, z)$  of  $(G, \prec, w)$ .

1.  $M \leftarrow$  a maximum-weight stable matching of  $(H, \prec, w|_{E(H)})$ ,  
 $(y, z) \leftarrow$  an integral optimal solution to  $\mathbb{D}(H, \prec, w|_{E(H)})$  (see Lemma 3.4),  
 $V^1 \leftarrow$  the set of all non-isolated vertices in  $H$ ,  $G_0 \leftarrow G$ ,  $i \leftarrow 0$
  2. **while**  $E(G_i) - E(H) \neq \emptyset$  **do**
  3. Find three distinct vertices  $r_i, s_i, t_i \in V(G_i)$  such that  $f_{G_i}(r_i) = s_i \in V^1$ ,  $l_{G_i}(s_i) = t_i$ ,  
and  $s_i t_i \in E(G_i) - E(H)$
  4.  $G_{i+1} \leftarrow G_i \setminus s_i t_i$ ,  $i \leftarrow i + 1$
  5. **endwhile**
  6. **for**  $j = i - 1$  **down to** 0 **do**
  7.  $r \leftarrow r_j$ ,  $s \leftarrow s_j$ ,  $t \leftarrow t_j$ ,  $\epsilon \leftarrow \max\{0, w(st) - y(s) - y(t) + z(\{e \in \delta_{j+1}(t) : st \prec_t e\})\}$ ,  
 $y(s) \leftarrow y(s) + \epsilon$ ,  $z(rs) \leftarrow z(rs) + \epsilon$ ,  $z(st) \leftarrow 0$
  8. **endfor**
  9. Output  $M$  and  $(y, z)$
-



**Lemma 3.6** *Let  $(G, \prec)$  be a preference system with a bipartite representation  $(H, \prec)$  and with  $G = (V, E)$ . Then for any  $w \in \mathbb{Z}^E$ , Algorithm 3.5 correctly finds a maximum-weight stable matching and a minimum-cost stable cover of  $(G, \prec, w)$  in  $O(m^2 \log m)$  time, where  $m = |E|$ .*

**Proof.** By Lemma 3.4, the stable matching  $M$  specified in line 1 can be found in  $O(m^2 \log m)$  time. From Lemma 2.2(i), we conclude that  $M$  is a maximum-weight stable matching of  $(G, \prec, w)$  as well.

Set  $k = |E - E(H)|$ . Then  $G_k = H$  and  $G_j$  is obtained from  $G_{j+1}$  by adding the edge  $s_j t_j$  for  $j = k - 1, k - 2, \dots, 0$ . Let  $(y_j, z_j)$  be the  $(y, z)$  generated in the algorithm corresponding to  $G_j$  for  $j = k, k - 1, \dots, 0$ . By line 1,  $(y_k, z_k)$  is an integral optimal solution to  $\mathbb{D}(G_k, \prec, w)$ . Thus, from the proof of Lemma 3.2 (with  $G_j$  in place of  $G$  and  $G_{j+1}$  in place of  $G'$ ), we deduce that  $(y_j, z_j)$  is an integral optimal solution to  $\mathbb{D}(G_j, \prec, w|_{E(G_j)})$  for  $j = k - 1, k - 2, \dots, 0$ . So the  $(y, z)$  output by the algorithm is an integral optimal solution to  $\mathbb{D}(G, \prec, w)$ .

It is easy to see that the complexity of the algorithm is dominated by line 1. Hence the whole algorithm runs in  $O(m^2 \log m)$  time.  $\blacksquare$

Now we are ready to establish the main result of this paper.

**Proof of Theorem 1.5.** From the Edmonds-Giles theorem stated in Section 1, we see that (iv) $\Rightarrow$ (iii). The implications (iii) $\Rightarrow$ (ii), (ii) $\Rightarrow$ (i), and (i) $\Rightarrow$ (iv) are established by Lemma 3.1, Lemma 2.3(iv), and Lemma 3.2, respectively. So the statements (i)-(iv) are equivalent. The second half of our theorem follows instantly from Lemma 3.6.  $\blacksquare$

By Lemma 3.1, if  $(G, \prec) \notin \mathfrak{S}$ , then  $FSM(G, \prec) - SM(G, \prec) \neq \emptyset$ . Using Theorem 1.5, we can explicitly find an element of this set.

**Corollary 3.7** *Let  $(G, \prec)$  be a preference system outside  $\mathfrak{S}$ , let  $(H, \prec)$  be its representation, and let  $\mathcal{C}$  be the semistable partition as defined in Lemma 2.3(iii). Then (recall (1.9))*

$$x_{\mathcal{C}} \in FSM(G, \prec) - SM(G, \prec).$$

**Proof.** From the proof of Lemma 3.1, we see that if  $x_{\mathcal{C}} \in SM(G, \prec)$ , then  $G_{\mathcal{C}}$  is a bipartite graph and hence so is  $H$  by Lemma 2.3(iv). It follows from Theorem 1.5 that  $(G, \prec) \in \mathfrak{S}$ ; this contradiction establishes the assertion.  $\blacksquare$

We finally arrive to statements concerning general preference systems.

**Proof of Theorem 1.7.** To establish the assertion, let us first show that

(1) for any  $w \in \mathbb{Z}^E$ , the problem  $\mathbb{D}(G, \prec, w)$  has a half-integral optimal solution  $(\bar{y}, \bar{z})$ , which can be found in  $O(m^2 \log m)$  time.

For this purpose, we construct a bipartite preference system  $(G', \prec')$ , with  $G' = (V', E')$ , and define a weight function  $w' \in \mathbb{Z}^{E'}$  as follows:

- each vertex  $v \in V$  corresponds to two vertices  $\bar{v}, \hat{v} \in V'$ ;

- each edge  $uv \in E$  corresponds to two edges  $\bar{u}\hat{v}, \hat{u}\bar{v} \in E'$  with weight  $w'(\bar{u}\hat{v}) = w'(\hat{u}\bar{v}) = w(uv)$ ; and
- each pair of adjacent edges  $vp, vq \in E$  with  $vp \prec_v vq$  in  $(G, \prec)$  corresponds to two pairs of adjacent edges  $\bar{v}\hat{p}, \hat{v}\bar{q} \in E'$  and  $\hat{v}\bar{p}, \bar{v}\hat{q} \in E'$  with  $\bar{v}\hat{p} \prec'_{\bar{v}} \hat{v}\bar{q}$  and  $\hat{v}\bar{p} \prec'_{\hat{v}} \bar{v}\hat{q}$  in  $(G', \prec')$ .

This completes the construction of  $(G', \prec')$  and definition of  $w'$ . We point out that this proof technique was first introduced by Abeledo and Rothblum in their proof of Theorem 3.3 in [1].

Let  $(y_*, z_*)$  be an optimal solution to  $\mathbb{D}(G, \prec, w)$ . Set  $y'(\bar{v}) = y'(\hat{v}) = y_*(v)$  for all  $v \in V$ , and set  $z'(\bar{u}\hat{v}) = z'(\hat{u}\bar{v}) = z_*(uv)$  for all  $uv \in E$ . It is a routine matter to check that  $(y', z')$  is a feasible solution to  $\mathbb{D}(G', \prec', w')$ , with objective value  $y'(V') - z'(E') = 2(y_*(V) - z_*(E))$ . So

(2) the optimal objective value of  $\mathbb{D}(G', \prec', w')$  is at most  $2(y_*(V) - z_*(E))$ .

Since  $G'$  is a bipartite graph with  $|E'| = 2m$ , by Theorem 1.3, we can find an integral optimal solution  $(y'_*, z'_*)$  to  $\mathbb{D}(G', \prec', w')$  in  $O(m^2 \log m)$  time. Using (2), we obtain

(3)  $y'_*(V') - z'_*(E') \leq 2(y_*(V) - z_*(E))$ .

Set  $\bar{y}(v) = (y'_*(\bar{v}) + y'_*(\hat{v}))/2$  for all  $v \in V$ , and set  $\bar{z}(uv) = (z'_*(\bar{u}\hat{v}) + z'_*(\hat{u}\bar{v}))/2$  for all  $uv \in E$ . It is easy to see that  $(\bar{y}, \bar{z})$  is a feasible solution to  $\mathbb{D}(G, \prec, w)$ , with objective value

(4)  $\bar{y}(V) - \bar{z}(E) = (y'_*(V') - z'_*(E'))/2 \leq y_*(V) - z_*(E)$  by (3).

From the optimality of  $(y_*, z_*)$ , we deduce that  $(\bar{y}, \bar{z})$  is also an optimal solution to  $\mathbb{D}(G, \prec, w)$ . Clearly,  $(\bar{y}, \bar{z})$  is half-integral and can be generated in  $O(m^2 \log m)$  time. Thus (1) holds, which implies that  $\pi(G, \prec)$  is totally dual half-integral.

By Lemma 3.4, an integral optimal solution  $x'_*$  to  $\mathbb{P}(G', \prec', w')$  can be found in  $O(m^2 \log m)$  time. Observe that

(5)  $(w')^T x'_* = y'_*(V') - z'_*(E')$ .

Set  $\bar{x}(uv) = (x'_*(\bar{u}\hat{v}) + x'_*(\hat{u}\bar{v}))/2$  for all  $uv \in E$ . It is straightforward to verify that  $\bar{x}$  is a feasible solution to  $\mathbb{P}(G, \prec, w)$ , with objective value  $w^T \bar{x} = (w')^T x'_*/2 = \bar{y}(V) - \bar{z}(E)$  by (5) and (4). So, from the LP duality theorem, we can conclude that  $\bar{x}$  is an optimal solution to  $\mathbb{P}(G, \prec, w)$ . Clearly,  $\bar{x}$  is half-integral and can be obtained in  $O(m^2 \log m)$  time.  $\blacksquare$

**Acknowledgments.** We are indebted to Professor Tamás Király and one anonymous referee for their invaluable comments and suggestions which led to substantial improvement on the presentation of our paper. Our special thanks also go to Professor Király for discovering a quick proof of the implication (i) $\Rightarrow$ (iv) in Theorem 1.5 (see Lemma 3.3 and the first proof of Lemma 3.2).

## References

- [1] H. Abeledo and U. Rothblum, Stable matchings and linear inequalities, *Discrete Appl. Math.* **54** (1994), 1-27.
- [2] J. Edmonds and R. Giles, A min-max relation for submodular functions on graphs, *Ann. Discrete Math.* **1** (1977), 185-204.

- [3] T. Feder, A new fixed point approach for stable networks and stable marriages, *J. Comput. System Sci.* **45** (1992), 233-284.
- [4] D. Gale and L. Shapley, College admissions and the stability of marriage, *Amer. Math. Monthly* **69** (1962), 9-15.
- [5] A. Goldberg and R. Tarjan, A new approach to the maximum-flow problem, *J. ACM* **35** (1988), 921-940.
- [6] D. Gusfield and R. Irving, *The Stable Marriage Problem: Structure and Algorithms*, MIT Press, Cambridge, 1989.
- [7] R. Irving, An efficient algorithm for the stable roommates problems, *J. Algorithms* **6** (1985), 577-595.
- [8] T. Király and J. Pap, Total dual integrality of Rothblum's description of the stable-marriage polyhedron, *Math. Oper. Res.* **33** (2008), 283-290.
- [9] D. Knuth, *Marriages stables et leurs relations avec d'autres problèmes combinatoires*, Les Presses de l'Université de Montreal, Montreal, 1976.
- [10] A. Roth and M. Sotomayor, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Cambridge University Press, Cambridge, 1990.
- [11] A. Roth, U. Rothblum, and J. Vande Vate, Stable matching, optimal assignments and linear programming, *Math. Oper. Res.* **18** (1993), 808-828.
- [12] U. Rothblum, Characterization of stable matchings as extreme points of a polytope, *Math. Programming* **54** (1992), 57-67.
- [13] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*, Springer-Verlag, Berlin, 2003.
- [14] J. Tan, A necessary and sufficient condition for the existence of a complete stable matching, *J. Algorithms* **12** (1991), 154-178.
- [15] C. Teo and J. Sethuraman, Geometry of fractional stable matchings and its applications, *Math. Oper. Res.* **23** (1998), 874-891.
- [16] J. Vande Vate, Linear programming brings marital bliss, *Oper. Res. Lett.* **8** (1989), 147-153.