



# Approximating Stable Matchings with Ties of Bounded Size

Jochen Koenemann<sup>1</sup>, Kanstantsin Pashkovich<sup>2</sup>, and Natig Tofigzade<sup>1(✉)</sup>

<sup>1</sup> University of Waterloo, Waterloo, ON N2L 3G1, Canada  
{jochen,natig.tofigzade}@uwaterloo.ca

<sup>2</sup> University of Ottawa, Ottawa, ON K1N 6N5, Canada  
kpashkov@uottawa.ca

**Abstract.** Finding a stable matching is one of the central problems in algorithmic game theory. If participants are allowed to have ties and incomplete lists, computing a stable matching of maximum cardinality is known to be NP-hard. In this paper we present a  $(3L - 2)/(2L - 1)$ -approximation algorithm for the stable matching problem with ties of size at most  $L$  and incomplete preferences. Our result matches the known lower bound on the integrality gap for the associated LP formulation.

**Keywords:** Stable matching · Approximation algorithms · Combinatorial optimization

## 1 Introduction

In an instance of the classical stable matching problem we are given a (complete) bipartite graph  $G = (A \cup B, E)$  where, following standard terminology, the nodes in  $A$  will be referred to as *men*, and the nodes in  $B$  represent *women*. Each man  $a \in A$  possesses a (strict, and complete) preference order over women in  $B$ , and similarly, all women in  $B$  have a preference order over men in  $A$ . A matching  $M$  in  $G$  is called *stable* if there are no *blocking pairs*  $(a, b)$ ; i.e. there do not exist  $(a, b) \notin M$  where both  $a$  and  $b$  prefer each other over their current partners in  $M$  (if there are any). In their celebrated work [4], Gale and Shapley proposed an efficient algorithm for finding a stable matching, providing a constructive proof that stable matchings *always* exist.

Stable matchings have wide-spread applications (e.g., see Manlove [3]), and many of these are large-scale. Therefore, as McDermid [15] points out, assuming that preferences are complete and strict is not realistic. Thus, in this paper, we will focus on stable matchings in the setting where preference lists are allowed to be incomplete and contain ties. Here, a woman is allowed to be indifferent between various men, and similarly, a man may be indifferent between several women. In this setting we consider the *maximum-cardinality stable matching* problem where the goal is to find a stable matching of maximum cardinality.

It is well-known that, in the settings where  $G$  is either complete or preferences do not contain ties, all stable matchings have the same cardinality [5].

Moreover, a straightforward extension of the algorithm in [4] solves our problem in these cases. When ties and incomplete preferences are permitted simultaneously, on the other hand, the problem of finding a maximum-cardinality stable matching is well-known to be NP-hard [14]. Furthermore, Yanagisawa [17] showed that it is NP-hard to find a  $(33/29 - \varepsilon)$ -approximate, maximum-cardinality stable matching. The same author also showed that assuming the *unique games* conjecture (UGC) it is hard to achieve performance guarantee of  $4/3 - \varepsilon$ .

On the positive side, maximum-cardinality stable matchings with ties and incomplete preferences have attracted significant attention [1, 6–12, 15, 16]. The best-known approximation algorithms for the problem achieve an approximation ratio of  $3/2$  [9, 15, 16].

How does the hardness of maximum-cardinality stable matching depend on the maximum allowed size of ties in the given instance? Huang and Kavitha [6] recently considered the case where the *size of any tie* is bounded by  $L = 2$ . The authors proposed an algorithm and showed that its performance guarantee is at most  $10/7$ . Chiang and Pashkovich [2] later provided an improved analysis for the same algorithm, showing that its real performance ratio is at most  $4/3$ , and this result is tight under the UGC [17]. Lam and Plaxton [13] very recently designed a  $1 + (1 - 1/L)^L$ -approximation algorithm for the so-called *one-sided* special case of our problem, where only preferences of men are allowed to have ties.

## 1.1 Our Contribution

Our main result is captured in the following theorem. Note that the integrality gap of the natural LP relaxation for the problem is at least  $(3L - 2)/(2L - 1)$  [8]. Hence, the performance ratio of our algorithm matches the known lower bound on the integrality gap.

**Theorem 1.** *Given an instance of the maximum-cardinality stable matching problem with incomplete preferences, and ties of size at most  $L$ ; the polynomial-time algorithm described in Sect. 2 finds a stable matching  $M$  with*

$$|M| \geq \frac{2L - 1}{3L - 2} |OPT|,$$

where  $OPT$  is an optimal stable matching.

Our algorithm is an extension of that by Huang and Kavitha [6] for ties of size two: every man has  $L$  proposals where each proposal goes to the acceptable women. Women can *accept* or *reject* these proposals under the condition that no woman holds more than  $L$  proposals at any point during the algorithm. Similar to the algorithm in [6], we use the concept of *promotion* introduced by Király [9] to grant men repeat chances in proposing to women. In comparison to [6], the larger number of proposals in our algorithm leads to subtle changes to

the forward and rejection mechanisms of women, and to further modifications to the way we obtain the output matching.

Our analysis is inspired by the analyses of both, Chiang and Pashkovich [2], and Huang and Kavitha [6], but requires several new ideas to extend it to the setting with larger ties. In both [6] and [2], the analyses are based on *charging schemes*: some objects are first assigned some values, called charges, and then charges are redistributed to nodes by a cost function. After a charging scheme is determined, relations between the generated total charges, and the sizes of output and optimal matchings are established, respectively, that lead to an approximation ratio. The analysis in [6] employs a complex charging scheme that acts *globally*, possibly distributing charges over the entire graph. In contrast, the charging scheme in [2] is *local* in nature, and exploits only the local structure of the output and optimal matchings, respectively.

We do not know of a direct way to extend the local *cost*-based analysis of [2] to obtain an approximation algorithm whose performance beats the best known  $3/2$ -approximation for the general case. Indeed we believe that any such improvement must involve a non-trivial change in the charging scheme employed. As a result, we propose a new analysis that combines local and global aspects from [2, 6]. The central technical novelty in the analysis is captured by Lemma 4 that provides an improved lower bound on the *cost* of components. As we will see later, our new charging scheme allows for a more fine-grained accounting of augmenting paths for the output matching of our algorithm.

## 2 Algorithm for Two-Sided Ties of Size Up to $L$

We introduce some notational conventions. Let  $a', a'' \in A$  be on the preference list of  $b \in B$ . We write  $a' \simeq_b a''$  if  $b$  is *indifferent* between  $a'$  and  $a''$ , and we write  $a' >_b a''$ , or  $a' \geq_b a''$  if  $b$  *strongly*, or *weakly* prefers  $a'$  over  $a''$ , respectively. The preferences of men over women are defined analogously. For  $c \in A \cup B$ , we let  $N(c)$  denote the set of nodes adjacent to  $c$  in  $G$ .

### 2.1 How Men Propose

Each man  $a \in A$  has  $L$  proposals  $p_a^1, p_a^2, \dots, p_a^L$ . A man starts out as *basic*, and later becomes *1-promoted* before he is eventually elevated to *2-promoted* status. Each man  $a \in A$  has a *rejection history*  $R(a)$  which records the women who rejected a proposal from  $a$  during his current promotion status. Initially, we let  $R(a) = \emptyset$ , for all  $a \in A$ .

Each proposal  $p_a^i$  for  $a \in A$  and  $i = 1, 2, \dots, L$  goes to a woman in  $N(a) \setminus R(a)$  most preferred by  $a$ , and ties are broken arbitrarily. If a proposal  $p_a^i$  for  $a \in A$  and  $i = 1, 2, \dots, L$  is rejected by a woman  $b \in B$ ,  $b$  is added to the rejection history of  $a$ , and subsequently,  $p_a^i$  is sent to a most preferred remaining woman in  $N(a) \setminus R(a)$ .

Suppose now that  $R(a)$  becomes equal to  $N(a)$  for some man  $a \in A$ . If  $a$  is either basic or 1-promoted then  $a$ 's rejection history is cleared, and  $a$  is promoted. Otherwise, if  $a$  is already 2-promoted,  $a$  stops making proposals.

### 2.2 How Women Decide

Each woman  $b \in B$  can hold up to  $L$  proposals, and among these more than one can come from the same man. Whenever she holds less than  $L$  proposals, newly received proposals are automatically accepted. Otherwise,  $b$  first tries to *bounce* one of her proposals, and if that fails, she will try to *forward* one of her proposals. If  $b$  can neither bounce nor forward a proposal, then  $b$  rejects a proposal.

We continue describing the details. In the following, we let  $P(b)$  and  $A(b)$  denote the set of proposals held by  $b \in B$  at the current point, and the set of men corresponding to these, respectively. Suppose that  $|P(b)| = L$ , and that  $b$  receives a new proposal  $p_a^i$  for some  $a \in A$  and  $i = 1, \dots, L$ .

**Bounce Step.** If there is a man  $\alpha \in A(b) \cup \{a\}$  and a woman  $\beta \in B \setminus \{b\}$  such that  $\beta \simeq_\alpha b$ , and  $\beta$  currently holds less than  $L$  proposals, then we move one of  $\alpha$ 's proposals from  $b$  to  $\beta$ , and we call the bounce step *successful*.

**Forward Step.** If there is a man  $\alpha \in A(b) \cup \{a\}$  and a woman  $\beta \in B \setminus \{b\}$  such that  $\beta \simeq_\alpha b$ , at least two proposals from  $\alpha$  are present in  $P(b)$ , no proposal from  $\alpha$  is present in  $P(\beta)$  and  $\beta$  is not in  $R(\alpha)$ , then  $b$  *forwards* a proposal  $p_\alpha^j \in P(b) \cup \{p_a^i\}$  for some  $j = 1, \dots, L$  to  $\beta$  and the forward step is called *successful*. As a consequence of a successful forward step,  $\alpha$  makes the proposal  $p_\alpha^j$  to  $\beta$ .

We point out that bounce and forward steps do not lead to an update to the rejection history of an involved man. To describe the rejection step, we introduce the following notions. For a woman  $b \in B$ , a proposal  $p_{a'}^i$  is called *more desirable* than  $p_{a''}^{i''}$  for  $a', a'' \in A$  and  $i', i'' = 1, \dots, L$  if  $b$  strongly prefers  $a'$  to  $a''$ , or if  $b$  is indifferent between  $a'$  and  $a''$  and  $a'$  has higher promotion status than  $a''$ . A proposal  $p_{a'}^i \in P(b)$  is *least desirable* in  $P(b)$  if  $p_{a'}^i$  is not more desirable than any proposal in  $P(b)$ . Whenever  $b \in B$  receives a proposal  $p_a^i$ ,  $|P(b)| = L$ , and neither *bounce* nor *forward* steps are successful, we execute a rejection step.

**Rejection Step.** If there is unique *least desirable proposal* in  $P(b) \cup \{p_a^i\}$ , then  $b$  rejects that proposal. Otherwise, if there are more than one least desirable proposal in  $P(b)$ ,  $b$  rejects a proposal from a man with the largest number of least desirable proposals in  $P(b) \cup \{p_a^i\}$ . If there are several such men, then we break ties arbitrarily. Subsequently,  $b$  is added to the rejection history of the man whose proposal is rejected.

### 2.3 The Algorithm

An approximate maximum-cardinality stable matching for a given instance  $G = (A \cup B, E)$  is computed in two stages.

**Stage 1.** Please see Algorithm 1 for the pseudo code for Stage 1.

Men propose in an arbitrary order and women bounce, forward or reject proposals as described above. The first stage finishes, when for each man  $a \in A$ , one of the following two conditions is satisfied: all proposals of  $a$  are accepted;  $R(a)$  becomes equal to  $N(a)$  for the third time.

We represent the outcome of the first stage as a bipartite graph  $G' = (A \cup B, E')$  with the node set  $A \cup B$  and the edge set  $E'$ , where each edge  $(a, b) \in E'$  denotes a proposal from  $a$  held by  $b$  at the end of the first stage. Note that  $G'$  may be a multigraph in which an edge of the form  $(a, b)$  appears with multiplicity equal to the number of proposals that  $b$  holds from  $a$ . Clearly, each node  $u$  in  $G'$  has degree at most  $L$ , denoted by  $\deg_{G'}(u) \leq L$ , since every man has at most  $L$  proposals that may be accepted and every woman can hold at most  $L$  proposals at any point in the first stage.

**Stage 2.** We compute a maximum-cardinality matching  $M$  in  $G'$  such that all nodes of degree  $L$  in  $G'$  are matched. The existence of such matching is guaranteed by Lemma 1. The result of the second stage is such a matching  $M$ , that is the output of the algorithm.

**Lemma 1.** *There exists a matching in the graph  $G'$  such that all nodes of degree  $L$  in  $G'$  are matched. Moreover, there is such a matching  $M$ , where all nodes of degree  $L$  in  $G'$  are matched and we have*

$$|M| \geq |E'|/L.$$

*Proof.* Consider the graph  $G' = (A \cup B, E')$  and the following linear program

$$\begin{aligned} \max \quad & \sum_{e \in E'} x_e \\ \text{s.t.} \quad & \sum_{e \in \delta(u)} x_e \leq 1 \quad (u \in A \cup B) \\ & \sum_{e \in \delta(u)} x_e = 1 \quad (u \in A \cup B, \deg_{G'}(u) = L) \\ & x \geq 0. \end{aligned}$$

It is well-known that the feasible region of the above LP is an integral polyhedron. Moreover, the above LP is feasible as is easily seen by considering the point that assigns  $1/L$  to each edge in  $E'$ . Hence there exists an integral point optimal for this linear program. Notice, that every integral point feasible for this linear program is a characteristic vector of a matching in  $G'$ , which matches all nodes of degree  $L$  in  $G'$ . To finish the proof, notice that the value of the objective function calculated at  $x^*$  equals  $|E'|/L$ . Thus the value of this linear program is at least  $|E'|/L$ , finishing the proof.  $\square$

## 2.4 Stability of Output Matching

Let the above algorithm terminate with a matching  $M$ . We first argue that it is stable.

---

**Algorithm 1.** Pseudo code for Stage 1 of the algorithm
 

---

```

1: let  $G = (A \cup B, E)$  be an instance graph, and  $N(c)$  denote the set of nodes adjacent
   to  $c \in A \cup B$  in  $G$ 
2: let  $G' = (A \cup B, E')$  be a multigraph with  $E'$  initialized to the empty multiset of
   edges
3: let  $\deg_{G'}(u)$  denote the degree of node  $u$  in  $G'$ , and  $A(b)$  denote the set of nodes
   adjacent to  $b \in B$  in  $G'$ 
4: for all  $a \in A$  do
5:      $R(a) := \emptyset$  ▷  $R(a)$  is the rejection history of man  $a$ 
6:      $stat_a := 0$  ▷  $stat_a$  is the promotion status of man  $a$ 
7: end for
8: while  $\exists a \in A$  s.t.  $\deg_{G'}(a) < L$  and  $R(a) \neq N(a)$  do
9:     let  $b \in N(a) \setminus R(a)$  be a woman s.t.  $b \succeq_a b'$  for all  $b' \in N(a) \setminus R(a)$ 
10:    PROPOSE( $a, b$ )
11: end while
12: return  $E'$ 
    {The following subroutine describes how  $b$  accepts the proposal from  $a$ , or bounces,
    forwards, or rejects a proposal}
13: procedure PROPOSE( $a, b$ )
14:    if  $\deg_{G'}(b) < L$  then
15:         $E' := E' \cup \{(a, b)\}$ 
16:    else if  $\exists \alpha \in A(b) \cup \{a\}$  and  $\exists \beta \in N(\alpha)$  s.t.  $\beta \simeq_\alpha b$  and  $\deg_{G'} \beta < L$  then
17:         $E' := E' \cup \{(a, b), (\alpha, \beta)\} \setminus \{(\alpha, b)\}$  ▷ bounce
18:    else if  $\exists \alpha \in A(b) \cup \{a\}$  and  $\exists \beta \in N(\alpha) \setminus R(\alpha)$  s.t.  $\beta \simeq_\alpha b$ ,
     $|(E' \cup \{(a, b)\}) \cap \{(\alpha, b)\}| \geq 2$  and  $\alpha \notin A(\beta)$  then
19:         $E' := E' \cup \{(a, b)\} \setminus \{(\alpha, b)\}$ 
20:        PROPOSE( $\alpha, \beta$ ) ▷ forward
21:    else
22:        let  $\mathcal{A}$  denote  $\{\alpha \in A(b) \cup \{a\} : \text{for all } a' \in A(b) \cup \{a\}, \alpha \leq_b a' \text{ and}$ 
         $\text{if } \alpha \simeq_b a', \text{ then } stat_\alpha \leq stat_{a'}\}$ 
23:        let  $\alpha_0$  be a man in  $\arg \max_{\alpha \in \mathcal{A}} |(E' \cup \{(a, b)\}) \cap \{(\alpha, b)\}|$ 
24:         $E' := E' \cup \{(a, b)\} \setminus \{(\alpha_0, b)\}$  ▷ reject
25:         $R(\alpha_0) := R(\alpha_0) \cup \{b\}$ 
26:        if  $R(\alpha_0) = N(\alpha_0)$  then
27:            if  $stat_{\alpha_0} < 2$  then
28:                 $stat_{\alpha_0} := stat_{\alpha_0} + 1$ 
29:                 $R(\alpha_0) := \emptyset$ 
30:            end if
31:        end if
32:    end if
33: end procedure
    
```

---

**Lemma 2.** *The output matching  $M$  is stable in  $G = (A \cup B, E)$ .*

*Proof.* Suppose for contradiction that  $M$  is not stable, i.e. suppose that there exists an edge  $(a, b) \in E$  that blocks  $M$ . If  $b$  rejected a proposal from  $a$  during the algorithm, then  $b$  holds  $L$  proposals when the algorithm terminates and all these proposals are from men, who are weakly preferred by  $b$  over  $a$ . Thus the

degree of  $b$  in  $G'$  is  $L$  implying that  $b$  is matched in  $M$  with a man, who is not less preferred by  $b$  than  $a$ . We get a contradiction to the statement that  $(a, b)$  blocks  $M$ .

Conversely, if  $b$  did not reject any proposal from  $a$  during the algorithm, then the algorithm terminates with all  $L$  proposals of  $a$  being accepted, particularly, by women, who are weakly preferred by  $a$  over  $b$ . Therefore the degree of  $a$  in  $G'$  is  $L$  implying that  $a$  is matched in  $M$  with a woman, who is not less preferred by  $a$  than  $b$ . Again, we get a contradiction to the statement that  $(a, b)$  is a blocking pair for  $M$ .  $\square$

## 2.5 Running Time

We now show that each stage of the algorithm has polynomial execution time. For the first stage, we illustrate that only a polynomial number of proposals are bounced, forwarded, or rejected during this stage. For the second stage, the proof of Lemma 2 implies that it is sufficient to find an optimal extreme solution for a linear program of polynomial size.

First, we show that proposals are bounced only polynomially many times. For every  $b \in B$ , at most  $L$  proposals may be bounced to  $b$ . Indeed, with each proposal bounced to  $b$ , the number of proposals held by  $b$  increases; also, the number of proposals held by  $b$  never decreases or exceeds  $L$  during the algorithm. Hence at most  $L|B|$  proposals are bounced during the first stage.

Second, we illustrate that proposals are forwarded only polynomially many times. For each  $a \in A$ , promotion status of  $a$ , and  $b \in B$  such that  $(a, b) \in E$ , at most one proposal of  $a$  may be forwarded to  $b$ . To see this, let  $b'$  be a woman forwarding a proposal of  $a$  to  $b$ . Notice that  $b$  cannot bounce the proposal after  $b$  receives it because, otherwise,  $b'$  could bounce it by the transitivity of indifference. Observe also that  $b$  may forward a proposal from  $a$  only if she holds another proposal from him. Then it follows from the forward step that no woman can forward a proposal of  $a$  to  $b$  as long as  $b$  holds a proposal from him. If  $b$  rejects the proposal, then she is added to the rejection history of  $a$ , and so  $b$  does not receive any proposal from  $a$  unless the promotion status of  $a$  changes. Hence at most  $3|A||B|$  proposals are forwarded during the first stage.

Finally, for each  $a \in A$ , promotion status of  $a$ , and  $b \in B$  such that  $(a, b) \in E$ ,  $b$  may reject at most  $L$  proposals from  $a$ . Indeed,  $b$  holds at most  $L$  proposals at any point in time, and since  $b$  is added to the rejection history of  $a$  after she rejected him,  $b$  does not receive any proposal from  $a$  unless the promotion status of  $a$  changes. Hence at most  $3L|A||B|$  proposals are rejected during the first stage.

## 3 Tight Analysis

Recall that  $\text{OPT}$  is a maximum-cardinality stable matching in  $G$ , and let  $M$  be the output matching defined above. If  $a \in A$  is matched with  $b \in B$  in  $\text{OPT}$ , we write  $\text{OPT}(a) := b$  and  $\text{OPT}(b) := a$ . Similarly, we use the notations  $M(a) := b$

and  $M(b) := a$  when  $a \in A$  is matched with  $b \in B$  in  $M$ . Note that our analysis is based on graph  $G'$  and therefore all graph-related objects will assume  $G'$ .

**Definition 1.** *A man  $a \in A$  is called successful if the algorithm terminates with all of his  $L$  proposals being accepted. Likewise, a woman  $b$  is called successful if she holds  $L$  proposals when the algorithm stops. In other words, a person  $c \in A \cup B$  is successful if the degree of  $c$  in  $G'$  is  $L$ , and unsuccessful otherwise.*

**Definition 2.** *A woman is called popular if she rejected a proposal during the algorithm, and unpopular otherwise.*

Remarks 1 and 2 below directly follow from the algorithm and are consequences of the bouncing step, and the rejection step, respectively.

*Remark 1.* Let  $a \in A$  and  $b, b' \in B$  be such that  $b$  holds a proposal from  $a$  when the algorithm finishes,  $b'$  is unsuccessful, and  $b' \simeq_a b$ . Then  $b$  is unpopular.

*Proof.* Suppose for contradiction that  $b$  is popular. Then at some point she could not bounce or forward any one of her proposals and so she was to reject a proposal. This implies that after  $b$  became popular, whenever she received a new proposal that could be bounced, that proposal would immediately be bounced. But then, when the algorithm terminates,  $b$  holds a proposal from  $a$ , that could successfully be bounced to  $b'$ , a contradiction.  $\square$

*Remark 2.* Let  $a, a' \in A$  and  $b \in B$  be such that  $b$  holds at least two proposals from  $a$  when the algorithm finishes,  $b$  rejected a proposal from  $a'$  at some point,  $a$  is basic, and  $a' \simeq_b a$ . Then there is an edge  $(a', b)$  in  $G'$ .

*Proof.* Suppose for a contradiction that  $(a', b) \notin G'$  holds. Let  $t$  be the most recent point in time when  $b$  rejects a proposal from  $a'$ . Then it follows from the algorithm that, at  $t$ ,  $a'' \geq_b a'$  holds for all  $a'' \in A(b)$ . The rejection step also implies that, at  $t$ , there is no  $a'' \in A$  such that  $a' \simeq_b a''$ ,  $a''$  is basic, and  $b$  holds more than one proposal from  $a''$ . Moreover, the algorithm implies that, after  $t$ , whenever she receives a new proposal from a man  $a''$  such that  $a'' <_b a'$ , she will immediately reject it unless she successfully bounces or forwards it. Now, consider a point in time after  $t$  when there is a man  $a''$  such that  $a' \simeq_b a''$ ,  $b$  already holds a proposal from  $a''$ , and receives another proposal from  $a''$ . Then the rejection step implies that she will reject one of the proposals from  $a''$  unless she successfully bounces or forwards it. But then, when the algorithm terminates,  $b$  holds at least two proposals from  $a$ , a contradiction.

### 3.1 Analytical Techniques

In the following, we define *inputs*, *outputs*, and *costs* – notions that are central in the analysis of our charging scheme. Before we take a closer look at these notions and define them formally, let us discuss phenomena captured by them.

We use two different objects, inputs and outputs, to differentiate between two different viewpoints on proposals accepted when the algorithm ends. In



particular, inputs are associated with the viewpoint of women on the proposals whereas outputs are associated with the viewpoint of men. The choice of terms “inputs” and “outputs” is due to the analysis in [6] where the edges of  $G'$  are directed from men to women, and so each proposal becomes an “input” for the woman, and analogously becomes an “output” for the corresponding man.

Now we describe the ideas that motivated our definitions concerning outputs and inputs. Let  $M + \text{OPT}$  denote the multiset that contains the edges in  $M$  and the edges in  $\text{OPT}$ . To establish the approximation guarantee of our algorithm, we analyze each connected component in  $M + \text{OPT}$ . In order to show that  $M$ -augmenting paths in  $M + \text{OPT}$  do not lead to a large approximation guarantee, we introduce the notions of *bad* and *good inputs* as well as *bad* and *good outputs*. For example, a certain number of bad inputs and bad outputs are generated by the edges incident to the endpoints of an  $M$ -augmenting path in  $M + \text{OPT}$ . Indeed, as we will see later, if  $a_0 - b_0 - a_1 - \dots - a_k - b_k$  is an  $M$ -augmenting path in  $M + \text{OPT}$  of length  $2k + 1$ ,  $k \geq 2$  where  $a_0 \in A$ , then  $b_0$  has at least  $L - 2$  bad inputs and  $a_k$  has at least  $L - 2$  bad outputs. Then to show the approximation guarantee of  $(3L - 2)/(2L - 1)$ , we provide a way to obtain a lower bound on the number of bad inputs and bad outputs of men and women in each  $M$ -augmenting path; and later we provide an upper bound on the total number of bad inputs and bad outputs of all men and women.

To implement the above ideas, we use a charging scheme. Our charging scheme associates a cost with each man and each woman. These costs keep track of bad inputs and bad outputs: bad inputs lead to an increase of the corresponding woman’s cost and bad outputs lead to an increase of the corresponding man’s cost. We show that the total cost of all men and women is bounded above by  $2L|M|$ . On the other side, we provide a lower bound on the total cost by giving a lower bound on the cost of each connected component in  $M + \text{OPT}$ . These upper and lower bounds lead to the desired approximation guarantee of  $(3L - 2)/(2L - 1)$ .

### 3.2 Inputs and Outputs

In our analysis inputs and outputs are fundamental edge-related objects for our charging scheme. Each edge in  $G'$  generates a certain number of charges. For example, as we will see in Sect. 3.3, if an edge  $(a, b)$  in  $G'$  belongs either to  $M$  or to  $\text{OPT}$ , two charges are generated by  $(a, b)$  so that one is carried to node  $a$  and one is carried to node  $b$  by cost function. To define similar charging mechanisms for the remaining types of edges in  $G'$ , we first distinguish them as in the following definitions.

**Definition 3.** *Given an edge  $(a, b)$  in  $G'$ , we say that  $(a, b)$  is an output from  $a \in A$  and an input to  $b \in B$  if  $(a, b)$  is not in  $M + \text{OPT}$ .*

To illustrate how outputs and inputs are determined, for example, let  $(a, b) \in M$ ,  $a \in A$ ,  $b \in B$  and  $n_{(a,b)}$  be the number of edges of the form  $(a, b)$  in the multigraph  $G'$ , then the edge  $(a, b)$  gives rise to the following number  $s_{(a,b)}$  of inputs (and to the same number of outputs)

$$s_{(a,b)} := \begin{cases} n_{(a,b)} - 1 & \text{if } (a,b) \notin \text{OPT} \\ 0 & \text{if } n_{(a,b)} = 1 \\ n_{(a,b)} - 2 & \text{otherwise.} \end{cases}$$

**Definition 4.** An input  $(a,b)$  to  $b \in B$  is called a bad input if one of the following is true:

- $b$  is popular and  $a >_b \text{OPT}(b)$ .
- $b$  is popular,  $a \simeq_b \text{OPT}(b)$ , but  $\text{OPT}(b)$  is unsuccessful.
- $b$  is popular,  $a$  is 1-promoted,  $\text{OPT}(b)$  is successful and  $M(b) \simeq_b \text{OPT}(b) \simeq_b a$ .

An input  $(a,b)$  to  $b \in B$  is a good input if it is not a bad input. In other words, an input  $(a,b)$  to  $b \in B$  is a good input if one of the following is true:

- $b$  is unpopular.
- $b$  is popular and  $\text{OPT}(b) >_b a$ .
- $b$  is popular,  $a \simeq_b \text{OPT}(b)$ ,  $\text{OPT}(b)$  is successful and  $a$  is not 1-promoted.
- $b$  is popular,  $a \simeq_b \text{OPT}(b)$ ,  $\text{OPT}(b)$  is successful, but not  $M(b) \simeq_b \text{OPT}(b) \simeq_b a$ .

An output  $(a,b)$  from a man  $a$  is called a bad output if one of the following is true:

- $b$  is unpopular.
- $b$  is popular,  $b >_a \text{OPT}(a)$ ,  $a$  is 1-promoted, but not  $M(b) \simeq_b \text{OPT}(b) \simeq_b a$ .
- $b$  is popular,  $b >_a \text{OPT}(a)$  and  $a$  is basic.

An output from a man  $a$  is a good output if that is not a bad output. In other words, an output  $(a,b)$  from a man  $a \in A$  is a good output if one of the following is true:

- $b$  is popular and  $\text{OPT}(a) \geq_a b$ .
- $b$  is popular,  $b >_a \text{OPT}(a)$  and  $a$  is 2-promoted.
- $b$  is popular,  $b >_a \text{OPT}(a)$ ,  $a$  is 1-promoted and  $M(b) \simeq_b \text{OPT}(b) \simeq_b a$ .

**Lemma 3.** There is no edge which is both a bad input and a bad output.

*Proof.* Assume that an edge  $(a,b)$ ,  $a \in A$ ,  $b \in B$  is both a bad input to  $b$  and a bad output from  $a$ . First, consider the first case from the definition of a bad input. It trivially contradicts all the cases from the definition of a bad output. Second, consider the first case from the definition of a bad input and either the second or the third case from the definition of a bad output. Then the case (1) below is implied. Third, consider the second case from the definition of a bad input and either the second or the third case from the definition of a bad output. Then the case (2) below is implied. Finally, consider the third case from the definition of a bad input. It trivially contradicts both the second and the third case from the definition of a bad output. Thus one of the following cases is true:

1.  $a >_b \text{OPT}(b)$ ;  $b >_a \text{OPT}(a)$ .
2.  $a \simeq_b \text{OPT}(b)$ , and  $\text{OPT}(b)$  is unsuccessful;  $a$  is not 2-promoted.

In case (1), the edge  $(a, b)$  is a blocking pair for  $\text{OPT}$ , contradicting the stability of  $\text{OPT}$ .

In case (2), since  $\text{OPT}(b)$  is unsuccessful,  $\text{OPT}(b)$  was rejected by  $b$  as a 2-promoted man. On the other hand,  $a \simeq_b \text{OPT}(b)$ ,  $a$  is not 2-promoted, and  $b$  holds a proposal from  $a$  when the algorithm terminates, contradicting the rejection step.  $\square$

**Corollary 1.** *The number of good inputs is at least the number of bad outputs.*

*Proof.* Assume for a contradiction that the number of good inputs is smaller than the number of bad outputs. Then there is an edge in  $G'$  which is a bad output but not a good input. In other words, there is an edge in  $G'$  which is both a bad output and a bad input, contradicting Lemma 3.  $\square$

### 3.3 Cost

In our charging scheme, cost is a function that assigns charges, that originate from the edges, to the nodes. More specifically, the cost of a man  $a$  is obtained by counting the edges in  $G'$  incident to  $a$ , where bad outputs contribute 2 and all other edges contribute 1. Similarly, the cost of a woman  $b$  is obtained by counting the edges in  $G'$  incident to  $b$ , to which good inputs contribute 0 and all other edges contribute 1.

In the following, let  $\text{deg}(u)$  be the degree of the node  $u$  in  $G'$ . For  $a \in A$ , we define his *cost* as follows:

$$\text{cost}(a) := \text{deg}(a) + k, \quad \text{where } k \text{ is the number of bad outputs from } a;$$

for  $b \in B$ , we define her cost as follows:

$$\text{cost}(b) := \text{deg}(b) - k, \quad \text{where } k \text{ is the number of good inputs to } b,$$

For a node set  $S \subseteq A \cup B$ ,  $\text{cost}(S)$  is defined as the sum of costs of all the nodes in  $S$ .

The above definitions lead to next three remarks.

*Remark 3.* Let  $b \in B$  be matched in  $M$  and have at least  $k$  bad inputs. Then  $\text{cost}(b) \geq k + 1$ .

*Proof.* Let  $k'$  be the number of good inputs to  $b$ . Since  $b$  is matched in  $M$ , the edge  $(M(b), b)$  is contained in  $G'$  and therefore it is not an input to  $b$ . Thus  $\text{deg}(b) \geq k + k' + 1$ . Hence, by definition of cost,  $\text{cost}(b) = \text{deg}(b) - k' \geq k + 1$  holds.  $\square$

*Remark 4.* Let  $b \in B$  be matched in  $\text{OPT}$ , have at least  $k$  bad inputs, and  $(\text{OPT}(b), b) \in E'$  where  $E'$  is the edge set of  $G'$ . Then  $\text{cost}(b) \geq k + 1$ .

*Proof.* Let  $k'$  be the number of good inputs to  $b$ . Since the edge  $(\text{OPT}(b), b)$  is in  $G'$ , it is not an input to  $b$ . Thus  $\deg(b) \geq k + k' + 1$ . So, by definition of cost,  $\text{cost}(b) = \deg(b) - k' \geq k + 1$  holds.  $\square$

*Remark 5.* Let  $b \in B$  be matched in both  $\text{OPT}$  and  $M$ ,  $\text{OPT}(b) \neq M(b)$ , and  $(\text{OPT}(b), b) \in E'$  where  $E'$  is the edge set of  $G'$ . Then  $\text{cost}(b) \geq 2$ .

*Proof.* Let  $k$  and  $k'$  be the numbers of bad inputs and good inputs to  $b$ , respectively. Since the edges  $(\text{OPT}(b), b)$  and  $(M(b), b)$  are contained in  $G'$ , they are not inputs to  $b$ . Thus  $\deg(b) \geq k + k' + 2$ . So, by definition of cost,  $\text{cost}(b) = \deg(b) - k' \geq k + 2 \geq 2$  holds.  $\square$

### 3.4 The Approximation Ratio

Let  $\mathcal{C}(M + \text{OPT})$  denote the set of connected components in a graph induced by the edge set  $M + \text{OPT}$ . Lemma 4 below bounds the cost of  $M + \text{OPT}$ . For the proof of this lemma, we refer the reader to the full version of the paper on [arXiv](#)

**Lemma 4.**  $\sum_{C \in \mathcal{C}(M + \text{OPT})} \text{cost}(C) \geq (L + 1)|\text{OPT}| + (L - 2)(|\text{OPT}| - |M|)$ .

We are ready to prove our main theorem, and restate it here for completeness.

**Theorem 1.** *Given an instance of the maximum-cardinality stable matching problem with incomplete preferences, and ties of size at most  $L$ ; the polynomial-time algorithm described in Sect. 2 finds a stable matching  $M$  with*

$$|M| \geq \frac{2L - 1}{3L - 2} |\text{OPT}|,$$

where  $\text{OPT}$  is an optimal stable matching.

*Proof.* By Lemma 1, we have

$$|M| \geq \frac{|E'|}{L} = \sum_{u \in A \cup B} \frac{\deg(u)}{2L}.$$

By definition of cost and by Corollary 1, we obtain

$$\sum_{u \in A \cup B} \deg(u) \geq \text{cost}(A \cup B).$$

Combining the above inequalities, we get

$$2L|M| \geq \sum_{u \in A \cup B} \deg(u) \geq \text{cost}(A \cup B) = \sum_{C \in \mathcal{C}(M + \text{OPT})} \text{cost}(C),$$

By Lemma 4, we obtain

$$2L|M| \geq \sum_{C \in \mathcal{C}(M + \text{OPT})} \text{cost}(C) \geq (L + 1)|\text{OPT}| + (L - 2)(|\text{OPT}| - |M|).$$

By rearranging the terms, we obtain

$$2L|M| + (L - 2)|M| \geq (L + 1)|\text{OPT}| + (L - 2)|\text{OPT}|,$$

and so we obtain the desired inequality

$$(3L - 2)|M| \geq (2L - 1)|\text{OPT}|.$$

□

### 3.5 Tightness of the Analysis

The following example shows that the bound in Theorem 1 is tight.

| Men's preferences   | Women's preferences   |
|---|---|
| $a_0 : (b_0 \ b_1^\gamma \ \dots \ b_{L-1}^\gamma)$                           | $b_0 : (a_0 \ a_1^\beta \ \dots \ a_{L-1}^\beta)$                               |
| $a_1^\alpha : (b_1^\alpha \ b_1^\gamma \ \dots \ b_{L-1}^\gamma)$             | $b_1^\alpha : a_1^\alpha \ a_1^\beta \ \dots \ a_{L-1}^\beta$                   |
| $\vdots$  | $\vdots$  |
| $a_{L-1}^\alpha : (b_{L-1}^\alpha \ b_1^\gamma \ \dots \ b_{L-1}^\gamma)$     | $b_{L-1}^\alpha : a_{L-1}^\alpha \ a_1^\beta \ \dots \ a_{L-1}^\beta$           |
| $a_1^\beta : (b_0 \ b_1^\alpha \ \dots \ b_{L-1}^\alpha) \ b_1^\beta$         | $b_1^\beta : a_1^\beta$   |
| $\vdots$  | $\vdots$  |
| $a_{L-1}^\beta : (b_0 \ b_1^\alpha \ \dots \ b_{L-1}^\alpha) \ b_{L-1}^\beta$ | $b_{L-1}^\beta : a_{L-1}^\beta$   |
| $a_1^\gamma : b_1^\gamma$   | $b_1^\gamma : (a_0 \ a_1^\alpha \ \dots \ a_{L-1}^\alpha) \ a_1^\gamma$         |
| $\vdots$  | $\vdots$  |
| $a_{L-1}^\gamma : b_{L-1}^\gamma$   | $b_{L-1}^\gamma : (a_0 \ a_1^\alpha \ \dots \ a_{L-1}^\alpha) \ a_{L-1}^\gamma$ |

**Fig. 1.** An instance with ties of size at most  $L$ ,  $L \geq 2$  for which the algorithm outputs a stable matching  $M$  with  $|\text{OPT}|/|M| = (3L - 2)/(2L - 1)$

*Example 1.* In Fig. 1, the preference list of each individual is ordered from a most preferred person to a least preferred one, where individuals within parentheses are tied. For example,  $a_1^\beta$  is indifferent between all the women in his preference list except  $b_1^\beta$ , who is less preferred than the others.

It is straightforward to check that there exists a unique maximum-cardinality stable matching, namely  $\text{OPT} = \{(a_0, b_0)\} \cup \{(a_i^j, b_i^j) \mid i = 1, \dots, L - 1, j = \alpha, \beta, \gamma\}$ . We show that there exists an execution of the algorithm which outputs the matching  $M = \{(a_0, b_0)\} \cup \{(a_i^\alpha, b_i^\gamma) \mid i = 1, \dots, L - 1\} \cup \{(a_i^\beta, b_i^\alpha) \mid i = 1, \dots, L - 1\}$ , leading to the ratio  $|\text{OPT}|/|M| = (3L - 2)/(2L - 1)$ .

*Proof.* The following is an execution of the algorithm which leads either to the matching  $M$  or a matching with the size of  $M$ .

- $a_0$  makes one proposal to every woman in his list; the women accept.
- $a_i^\alpha$  for all  $i = 1, \dots, L - 1$  makes one proposal to every woman in his list; the women accept.
- $a_i^\beta$  for all  $i = 1, \dots, L - 1$  makes one proposal to every woman except the last one in his list; the women accept.
- $a_i^\gamma$  starts to propose  $b_i^\gamma$  for all  $i = 1, \dots, L - 1$ , but each time  $a_i^\gamma$  makes a proposal, the proposal is rejected;  $a_i^\gamma$  gives up.

□

## References

1. Bauckholt, F., Pashkovich, K., Sanità, L.: On the approximability of the stable marriage problem with one-sided ties. ArXiv e-prints (2018)
2. Chiang, R., Pashkovich, K.: On the approximability of the stable matching problem with ties of size two. *Algorithmica* 1–19 (2020). <https://doi.org/10.1007/s00453-020-00703-9>
3. David, M.: *Algorithmics of Matching Under Preferences*, vol. 2. World Scientific, Singapore (2013)
4. Gale, D., Shapley, L.: College admissions and the stability of marriage. *Am. Math. Monthly* **69**(1), 9–15 (1962)
5. Gale, D., Sotomayor, M.: Some remarks on the stable matching problem. *Discrete Appl. Math.* **11**(3), 223–232 (1985)
6. Huang, C.-C., Kavitha, T.: Improved approximation algorithms for two variants of the stable marriage problem with ties. *Math. Program.* 353–380 (2015). <https://doi.org/10.1007/s10107-015-0923-0>
7. Iwama, K., Miyazaki, S., Yamauchi, N.: A 1.875-approximation algorithm for the stable marriage problem. In: 18th Symposium on Discrete Algorithms (SODA), pp. 288–297 (2007)
8. Iwama, K., Miyazaki, S., Yanagisawa, H.: A 25/17-approximation algorithm for the stable marriage problem with one-sided ties. *Algorithmica* **68**(3), 758–775 (2014). <https://doi.org/10.1007/s00453-012-9699-2>
9. Király, Z.: Better and simpler approximation algorithms for the stable marriage problem. *Algorithmica* **60**(1), 3–20 (2011)
10. Király, Z.: Linear time local approximation algorithm for maximum stable marriage. *Algorithms* **6**(3), 471–484 (2013)
11. Lam, C.K.: *Algorithms for stable matching with indifferences*. Ph.D. thesis, University of Texas at Austin (2019)
12. Lam, C.K., Plaxton, C.G.: A  $(1 + 1/e)$ -approximation algorithm for maximum stable matching with one-sided ties and incomplete lists. In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2823–2840. SIAM, Philadelphia (2019). <https://doi.org/10.1137/1.9781611975482.175>
13. Lam, C.-K., Plaxton, C.G.: Maximum stable matching with one-sided ties of bounded length. In: Fotakis, D., Markakis, E. (eds.) *SAGT 2019*. LNCS, vol. 11801, pp. 343–356. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-30473-7\\_23](https://doi.org/10.1007/978-3-030-30473-7_23)
14. Manlove, D.F., Irving, R.W., Iwama, K., Miyazaki, S., Morita, Y.: Hard variants of stable marriage. *Theoret. Comput. Sci.* **276**(1–2), 261–279 (2002). [https://doi.org/10.1016/S0304-3975\(01\)00206-7](https://doi.org/10.1016/S0304-3975(01)00206-7)

15. McDermid, E.: A  $3/2$ -approximation algorithm for general stable marriage. In: Albers, S., Marchetti-Spaccamela, A., Matias, Y., Nikolettseas, S., Thomas, W. (eds.) ICALP 2009, Part I. LNCS, vol. 5555, pp. 689–700. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-642-02927-1\\_57](https://doi.org/10.1007/978-3-642-02927-1_57)
16. Paluch, K.: Faster and simpler approximation of stable matchings. *Algorithms* **7**(2), 189–202 (2014)
17. Yanagisawa, H.: Approximation algorithms for stable marriage problems. Ph.D. thesis, Kyoto University (2007)