Egerváry Research Group
on Combinatorial Optimization

Technical reportS

TR-2008-04. Published by the Egrerváry Research Group, Pázmány P. sétány 1/C, H–1117, Budapest, Hungary. Web site: www.cs.elte.hu/egres. ISSN 1587–4451.

Better and simpler approximation algorithms for the stable marriage problem

Zoltán Király

April 4, 2008
Revised on June 30, 2008

Better and simpler approximation algorithms for the stable marriage problem

Zoltán Király*

Abstract

We first consider the problem of finding a maximum stable matching if incomplete lists and ties are both allowed, but ties only for one gender. For this problem we give a simple, linear time 3/2-approximation algorithm, improving on the best known approximation factor 5/3 of Irving and Manlove [5]. Next, we show how this extends to the Hospitals/Residents problem with the same ratio if the residents have strict orders. We also give a simple linear time algorithm for the general problem with approximation factor 5/3, improving the best known 15/8-approximation algorithm of Iwama, Miyazaki and Yamauchi [7]. For the cases considered in this paper it is NP-hard to approximate within a factor of 21/19 by the result of Halldórsson et al. [3].

Our algorithms not only give better approximation ratios than the cited ones, but are much simpler and run significantly faster. Also we may drop a restriction used in [5] and the analysis is substantially more moderate.

Keywords: stable matching, Hospitals/Residents problem, approximation algorithms

1 Introduction

An instance of the stable marriage problem consists of a set U of N men, a set V of N women, and a preference list for each person, that is a weak linear order (ties are allowed) on some members of the opposite gender. A pair $(m \in U,\ w \in V)$ is called acceptable if m is on the list of w and w is on the list of m. We model acceptable pairs with a bipartite graph $G = (U, V, E)$, (where E is the set of acceptable pairs; we may assume that if w is not on the list of m then m is also missing from the list of w). A matching in this graph consists of mutually disjoint acceptable pairs. A matching M is *stable* if there is no blocking pair, where an acceptable pair is *blocking* if they strictly prefer each other to their current partners (the exact

*Department of Computer Science and Communication Networks Laboratory, Eötvös University, Pázmány Péter sétány 1/C Budapest, Hungary H-1117. Research is supported by EGRES group (MTA-ELTE) and OTKA grants NK 67867, K 60802 and T 046234, and by Hungarian National Office for Research and Technology programme NKFP072-TUDORKA7. E-mail: kiraly@cs.elte.hu

definition is given below). It is well-known that a stable matching always exists and can be found in linear time. An interesting problem, motivated by applications, is to find a stable matching of maximum size. This problem is known to be NP-hard for even very restricted cases [6, 8]. Moreover, it is APX-hard [2] and cannot be approximated within a factor of $21/19 - \delta$, even if ties occur only in the preference lists of one gender, furthermore if every list is either totally ordered or consists of a single tied pair [3]. As the applications of this problem are important, researchers started to develop good approximation algorithms in the last decade. We say that an algorithm is approximating with factor r if it gives a stable matching M with size $|M| \geq (1/r) \cdot |M_{\text{opt}}|$ where M_{opt} is a stable matching of maximum size. It is easy to give a 2-approximating algorithm, as any stable matching is maximal. The first non-trivial approximation algorithm was given by Halldórsson et al. [3] and was recently improved by Iwama, Miyazaki and Yamauchi [7] to a 15/8-approximation. This was later improved for the special case, where ties are allowed for only one gender and only at the ends of the lists, by Irving and Manlove [5]. (We must emphasize that the second restriction is not needed for our results.) They gave a 5/3-approximating algorithm for this special case. Their algorithm also applies for the Hospitals/Residents problem (see later) if residents have strictly ordered lists. If, moreover, ties are of size 2, Halldórsson et al. [3] gave an 8/5-approximation and in [4] they described a randomized algorithm for this special case with expected factor of 10/7. The paper of Irving and Manlove [5] also gives a detailed list of known and possible applications that motivate investigating approximation algorithms.

We store the lists as priorities. For an acceptable pair (m, w) let $\text{pri}(m, w)$ be an integer from 1 up to N representing the priority of w for m. We say that $m \in U$ strictly prefers $w \in V$ to $w' \in V$ if $\text{pri}(m, w) > \text{pri}(m, w')$. Ties are represented by the same number, e.g., if m equally prefers w_1, w_2 and w_3 then $\text{pri}(m, w_1) = \text{pri}(m, w_2) = \text{pri}(m, w_3)$. Of course, $\text{pri}(m, w)$ is not related to $\text{pri}(w, m)$. We represent these priorities by writing $\text{pri}(m, w)$ and $\text{pri}(w, m)$ close to the corresponding end of edge mw.

Let M be a matching. If m is *matched* in M, or in other words m is not *single*, we denote m's partner by $M(m)$. Similarly we use $M(w)$ for the partner of woman w. A pair (m, w) is *blocking* if $mw \in E \setminus M$ (they are an acceptable pair and they are not matched) and

- m is either single or $\text{pri}(m, w) > \text{pri}(m, M(m))$, and
- w is either single or $\text{pri}(w, m) > \text{pri}(w, M(w))$.

The famous algorithm of Gale and Shapley [1] for finding a stable matching is the following. Initially every man is active and makes any strict order of acceptable women according to the priorities (higher priority comes before lower).

Each active man m proposes to the next woman w on his strict list if w exists, otherwise (if he has processed the whole list) m inactivates himself. If the proposal was (temporarily) accepted then m inactivates himself, otherwise, if m was rejected, m keeps on proposing to the next woman from his list.

Each woman w who got some proposals keeps the best man as a partner and rejects all other men. More precisely, the first man m who proposed to w will be her first partner ($M(w) := m$). Later if w gets a new proposal from another man m', she rejects m' if $\text{pri}(w, m') \leq \text{pri}(w, M(w))$; otherwise w rejects $M(w)$, then $M(w)$ is re-activated, and finally w keeps $M(w) := m'$ as a new partner. The algorithm finishes if every man is inactive (either has a partner or has searched over his strict list). This algorithm runs in $O(|E|)$ time if G is given by edge-lists and sorting is done by bucket sort (we may suppose that every person has a non-empty list).

Theorem 1.1 (Gale-Shapley). *Algorithm GS defined above always ends in a stable matching M.*

Proof. Let $mw \in E \setminus M$. If m never made a proposal to w then in the end he has a partner w' who precedes w on m's strict list, consequently $\text{pri}(m, w') \geq \text{pri}(m, w)$. Otherwise, w rejected m at some point, when w had a partner m' not worse than m. Observe that after w received a proposal, she will always have a partner. Moreover, when w changes partner, she always chooses a (strictly) better one. Thus in the end $\text{pri}(w, M(w)) \geq \text{pri}(w, m') \geq \text{pri}(w, m)$, so mw is not blocking. □

In what follows, we will use not only the statement of this theorem (as most of the previous results do), but the Algorithm GS itself with some modifications/extensions.

In the *Hospitals/Residents* problem the roles of women are played by hospitals and the roles of men are played by residents. Moreover, each hospital w has a positive integer capacity $c(w)$ (the number of free positions). Instead of matchings we consider *assignments*, that is a subgraph F of G, such that all residents have degree at most one in F, and each hospital w has degree at most $c(w)$ in F. For a resident m who is assigned, $F(m)$ denotes the corresponding hospital. For a hospital w, $F(w)$ denotes the set of residents assigned to it. We say that hospital w is *full* if $|F(w)| = c(w)$ and otherwise *under-subscribed*. Here a pair (m, w) is *blocking* if $mw \in E \setminus F$ (they are an acceptable pair and they are not assigned to each other) and

- m is either single or $\text{pri}(m, w) > \text{pri}(m, F(m))$, and
- w is either under-subscribed or $\text{pri}(w, m) > \text{pri}(w, m')$ for at least one resident $m' \in F(w)$.

An assignment is *stable* if there is no blocking pair. It is easy to modify Algorithm GS to give a stable assignment for the Hospitals/Residents problem. Each hospital w manages to keep a set of buckets indexed by integers up to N, containing each assigned resident m in the bucket indexed by $\text{pri}(w, m)$; and w also stores the number of assigned residents and a pointer to the first non-empty bucket. If hospital w gets a new proposal from resident m then it accepts him either if w is under-subscribed or if $\text{pri}(w, m) > \text{pri}(w, m')$ for the worst assigned resident m'. Apart from this, the algorithm is the same. It clearly gives a stable assignment, and it is easy to see that also runs in $O(|E|)$ time (rejections can be decided in constant time as well as updating the data when accepting). We call this modified GS algorithm HRGS. As before, we are interested in giving a maximum size assignment, i.e., a stable assignment F with maximum number of edges (that is a maximum number of assigned residents).

In the next section we consider the special case of the maximum stable marriage problem, where each man's list is strictly ordered. We allow arbitrary number of arbitrarily long ties for each woman. We give a simple algorithm running in time $O(|E|)$. First we run Algorithm GS, then we give extra scores to single men, that raise their priorities. These men are re-activated and start making proposals from the beginning of their lists. A simple proof shows that this slightly modified algorithm gives a 3/2-approximation to the maximum stable marriage problem.

In Section 3 we show that this algorithm applies to the Hospitals/Residents problem as well in the (practically plausible) case when residents have strictly ordered lists, also giving 3/2-approximation for the maximum assignment in time $O(|E|)$.

Section 4 contains a slightly more complicated algorithm for the general case. First we run the algorithm of Section 2, then change the roles of men and women. In the second phase women get extra scores and make proposals to men. This algorithm still runs in linear-time, and gives a 5/3-approximation. Finally we propose some open problems.

2 Men have strictly ordered lists

In this section we suppose that the lists of men are strictly ordered. We are going to define extra scores, $\pi(m)$ for every man with the following properties. Initially $\pi(m) = 0$ and at any time $0 \leq \pi(m) < 1$ for each man. We also define adjusted priorities: $\text{pri}'(m, w) := \text{pri}(m, w)$ and $\text{pri}'(w, m) := \text{pri}(w, m) + \pi(m)$ for each acceptable pair (m, w). It is straightforward to see that if M is stable with respect to pri′ then it is also stable with respect to pri.

We define a modification of Algorithm GS, that is called rmGS (reduced men-proposal GS), as follows. This algorithm starts with a stable matching, given extra scores and a set of active men. Run the original GS algorithm (active men make proposals; at the beginning of the algorithm they start from the beginning of their strict lists), where women use pri′ to decide rejections. Stop when every man is inactive.

If some men with zero extra score remained single, we increase the score of those men to ε and re-activate them. In the next round they start making proposals from the beginning of their strict list. At any time let SM denote the set of single men, and $\Pi_0 := \{m \in U : \pi(m) = 0\}$. We fix $\varepsilon = 1/2$.

Our approximation algorithm is as follows:

ALGORITHM GSA1
run GS
FOR $m \in U \quad \pi(m) := 0$
WHILE $SM \cap \Pi_0 \neq \emptyset$
  FOR $m \in SM \cap \Pi_0$
    $\pi(m) := \varepsilon$
    re-activate m
  run rmGS

This simple algorithm runs in $O(|E|)$ time, as there are at most $2|E|$ proposals altogether. It is easy to see that Algorithm GSA1 gives a stable matching M with respect to the adjusted priority, hence M is stable for the original problem as well.

Let M_{opt} denote any maximum size stable matching (stable for the original priorities).

Theorem 2.1. *If men have strictly ordered preference lists, M is the output of Algorithm GSA1 and M_{opt} is a maximum size stable matching then*

$$|M_{\text{opt}}| \leq \frac{3}{2} \cdot |M|.$$

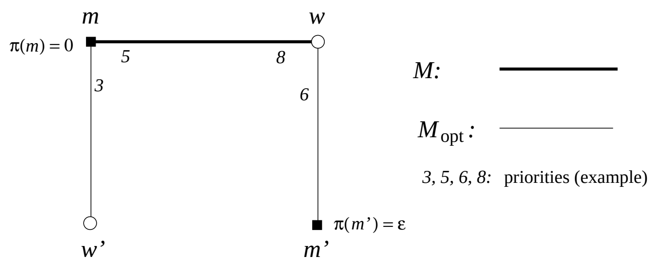

Figure 1: A path of length three in $M \cup M_{\text{opt}}$

Proof. We use an idea of Iwama, Miyazaki and Yamauchi [7]. Take the union of M and M_{opt}. We consider common edges as a two-cycle. Each component of $M \cup M_{\text{opt}}$ is either an alternating cycle (of even length) or an alternating path. It is enough to prove that in each component there are at most $3/2$ times as many M_{opt}-edges as M-edges. This is clearly true for each component except for alternating paths of length three with the M-edge mw in the middle (see Figure 1).

We claim that such a component cannot exist. Suppose that $M(m) = w$, $M_{\text{opt}}(m) = w' \neq w$, $M_{\text{opt}}(w) = m' \neq m$ and that m' and w' are single in M. Observe first that w' never got a proposal during Algorithm GSA1. Consequently $\pi(m) = 0$ at the end, as otherwise he would have proposed to each acceptable woman. We may also conclude that $\text{pri}(m, w) > \text{pri}(m, w')$ because there are no ties in the men's lists. When the algorithm finishes, $\pi(m') = \varepsilon$, and m' proposed to every acceptable woman with this extra score, but w rejected him. This means that $\text{pri}(w, m) = \text{pri}'(w, m) \geq \text{pri}'(w, m') = \text{pri}(w, m') + \varepsilon$ consequently $\text{pri}(w, m) > \text{pri}(w, m')$. However, in this case edge mw blocks M_{opt}, a contradiction. $\square$

We have an example (see Figure 2) showing that for our algorithm this bound is tight (a possible order of proposals and extra score increases is the following: $mw, m'w, m'w'', m''w''$, $\pi(m'') = \varepsilon$, $m''w''$).

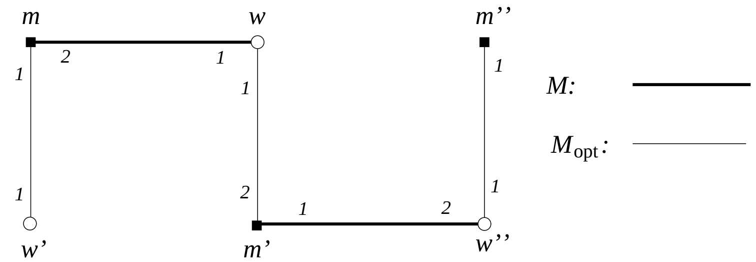

Figure 2: An example where GSA1 gives $|M| = (2/3) \cdot |M_{\text{opt}}|$

Note: for open questions please see the section "Open Problems".

3 Hospitals/Residents with strictly ordered residents' lists

We consider the Hospitals/Residents problem with the restriction that residents have strict orders on acceptable hospitals. Note, that for real-life applications of this scheme, this assumption is realistic. Here, as appropriate, residents get extra scores. The adjusted priorities are defined as in Section 2.

For a reader familiar with this topic it is straightforward that after "cloning" of hospitals the previous algorithm runs with the same approximation ratio. However, we describe an algorithm for this problem in some detail for not only to newcomers, but for three more reasons: (i) the cloning is not well defined in the literature, (ii) we give a linear time algorithm, and (iii) for showing an example and a theorem at the end of this section.

We modify GSA1 by replacing GS by HRGS and define rmHRGS as a modification of HRGS analogously to the derivation of rmGS from GS. Here SM denotes the set of unassigned residents and again $\Pi_0 := \{m \in U : \pi(m) = 0\}$.

Algorithm hrgsa1
run HRGS
FOR $m \in U \quad \pi(m) := 0$
WHILE $SM \cap \Pi_0 \neq \emptyset$
    FOR $m \in SM \cap \Pi_0$
        $\pi(m) := \varepsilon$
        re-activate m
    run rmHRGS

Algorithm HRGSA1 also runs in time $O(|E|)$ (hospitals need to have $2N$ buckets), and gives a stable assignment F.

Theorem 3.1. *If residents have strictly ordered preference lists, F is the output of Algorithm HRGSA1 and F_{opt} is any maximum size stable assignment then*

$$|F_{\text{opt}}| \le \frac{3}{2} \cdot |F|.$$

Proof. We suppose that positions at hospital w are numbered by $1 \dots c(w)$. For the proof we make an auxiliary bipartite graph $G' = (U, V', E')$ and new preference lists as follows. The set U of residents remains unchanged. The set V' consists of the positions, i.e., $V' = \{w^i : w \in V,\ 1 \le i \le c(w)\}$. An edge connects resident m and position w^i if (m, w) was an acceptable pair (if hospital w was acceptable to m then all positions at w are acceptable to m). Each position w^i inherits the preference list of hospital w. For resident m we have to make a new (and also strict) preference list. Take the original list, and replace each w by $w^1 < w^2 < \ldots < w^{c(w)}$ (thus if w_1 was preferred by m to w_2 then all positions of w_1 will be preferred to all positions of w_2). If F is an assignment in G then it defines a matching M in G' by distributing edges of F incident to a hospital w to distinct positions $w^1, w^2, \dots, w^{d_F(w)}$. And, conversely, any matching M of G' defines an assignment in G. The crucial observation is that if assignment F is stable in G then the associated matching M is stable in G', and if matching M is stable in G then the associated assignment F is stable in G. Moreover, if we imagine running Algorithm GSA1 on G', the resulting matching M corresponds to the assignment F given by Algorithm HRGSA1. Using these observations Theorem 2.1 implies this theorem. □

We note that the example on Figure 2 can be easily modified to show that this algorithm cannot achieve better approximation ratio than 3/2, not even if all hospitals have large capacities and if each hospital has an absolutely unordered list (i.e., $\text{pri}(w, m) = 1$ for every acceptable resident m).

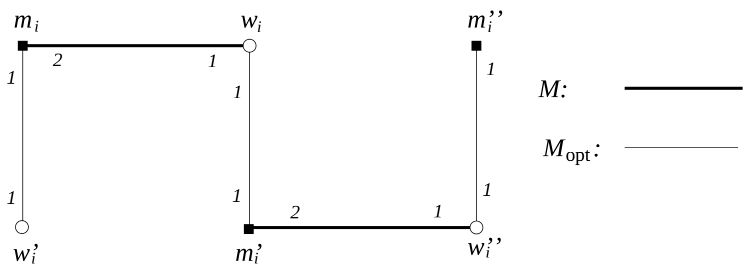

Figure 3: A building block of the example where HRGSA1 gives$|F| = (2/3) \cdot |F_{\text{opt}}|$

We make c copies of the example shown in Figure 3, one for each $i = 1 \dots c$. Then glue together the c copies of w_i, the c copies of w_i' and the c copies of w_i''. Assign capacity c to each hospital (w, w' and w''). The following is a possible run of Algorithm HRGSA1 yielding an assignment F with $|F| = 2c$, while $|F_{\text{opt}}| = 3c$. First every resident m_i'' proposes to hospital w''. Next, every resident m_i proposes to hospital w; now hospitals w and w'' are full. Then every resident m_i' proposes first to w'' and then to w, but they are always rejected. So every resident m_i' gets an extra score. They propose again to hospital w'' and they succeed. Now every resident m_i'' gets an extra score, and proposes again to w'' but they are rejected.

However, with a different type of restriction we are able to prove a stronger theorem. For a hospital w let $\tau(w)$ denote the length of the longest tie for w, and let $\lambda := \max_{w \in V} \tau(w)/(2c(w))$.

Theorem 3.2. *Algorithm HRGSA1 gives approximation ratio not worse than*

$$\frac{3}{2} - \frac{1}{6} \cdot \frac{1-\lambda}{1+\lambda}$$

Proof. The proof is very technical, so we only sketch the idea of it. Every component of $M \cup M_{\text{opt}}$ (in G') that is a 5-path has a middle hospital-position w^i such that hospital w is full. Each such hospital has at most $\tau(w)/2$ positions in such a bad component and $c(w) - \tau(w)/2 \geq \frac{1-\lambda}{2\lambda}\tau(w)$ other positions lying in a good component (where the ratio of F-edges against the F_{opt}-edges is at least 3/4). In the "worst case" this component is a 7-path that can contain at most 3 such hospital-positions. □

4 General stable marriage

Now we consider the general maximum stable marriage problem. First we run the algorithm of Section 2, then change the roles of men and women. In the second phase women get extra scores and propose to men.

Accordingly, we also use extra scores $\pi(w)$ for women: initially $\pi(w) = 0$ and at any time $0 \leq \pi(w) < 1$ for each woman w. We also re-define adjusted priorities: $\text{pri}'(m, w) := \text{pri}(m, w) + \pi(w)$ and $\text{pri}'(w, m) := \text{pri}(w, m) + \pi(m)$ for each acceptable pair (m, w). It is straightforward to see that if M is stable with respect to pri' then it is also stable with respect to pri.

In the first phase we run Algorithm GSA1, women do not get extra scores in this phase. Next, in the second phase we change the roles of men and women, in this phase we increase extra scores of women only. At the beginning of the second phase each woman makes any strict order of acceptable men according to the adjusted priorities (higher priority comes before lower).

We define Algorithm rwGS (reduced woman-proposal GS) similarly to Algorithm rmGS. The algorithm starts with a stable matching, given extra scores and a set of active women. Run the original GS algorithm with interchanged roles: active women make proposals, and men use pri' to decide rejections. But here we have a major difference. If a woman w with $\pi(w) = 0$ is rejected by her actual partner at any time during the process then she gets $\pi(w) := \varepsilon/2$ extra scores, activates herself, and starts making proposals *from the beginning of her strict list*. Stop when every woman is inactive.

If some women with less than ε extra score remained single, we increase the score of those women to ε and re-activate them. In the next round they start making proposals from the beginning of their strict list. At any time let SW denote the set of single women and $\Pi := \{w \in V : \pi(w) \leq \varepsilon/2\}$. We also use $\varepsilon = 1/2$.

Our approximation algorithm is as follows.

ALGORITHM GSA2
Phase 1
run GSA1
Phase 2
FOR $w \in V \quad \pi(w) := 0$
WHILE $SW \cap \Pi \neq \emptyset$
    FOR $w \in SW \cap \Pi$
        $\pi(w) := \varepsilon$
        re-activate w
    run rwGS

First we claim that the algorithm runs in time $O(|E|)$. To see this we must consider two things. In Phase 2, every woman processes her strict list at most twice, so there are at most $2|E|$ proposals in the second phase. The strict lists of women can be calculated in $O(|E|)$ time altogether using bucket sort with $2N$ buckets.

Lemma 4.1. *The matching M given by Algorithm GSA2 is stable with respect to* pri′ *consequently it is stable with respect to* pri.

Proof. We use the facts that in Phase 1 the positions of women do not decline, while during Phase 2 the positions of men do not decline. Let mw be any edge in $E \setminus M$. First suppose that at the end $\pi(w) > 0$. After woman w got her final extra score, she started to propose to men: either w did not propose to m, in this case $\text{pri}'(w, m) \leq \text{pri}'(w, M(w))$; or else w proposed to m but m rejected her, in this case $\text{pri}'(m, w) \leq \text{pri}'(m, M(m))$. In both cases we get that the edge mw is not blocking. Now suppose that at the end $\pi(w) = 0$. In this case w is matched in M, and also matched in M', where M' denotes the matching at the end of Phase 1. Moreover $M(w) = M'(w) = m' \neq m$. In Phase 1, after man m got his final score, either m did not propose to w, in this case $\text{pri}'(m, M(m)) \geq \text{pri}((m, M(m)) \geq \text{pri}(m, M'(m)) \geq \text{pri}(m, w) = \text{pri}'(m, w)$; or else m proposed to w but w rejected him, in this case $\text{pri}'(w, M(w)) = \text{pri}'(w, M'(w)) \geq \text{pri}'(w, m)$. In both cases we get again that the edge mw is not blocking. □

Theorem 4.2. *If M is the output of Algorithm GSA2 and M_{opt} is any maximum size stable matching then*

$$|M_{\text{opt}}| \leq \frac{5}{3} \cdot |M|.$$

Proof. First we need a technical lemma. Let M' denote the matching given at the end of Phase 1. Consider components of $M \cup M_{\text{opt}}$ as before.

Lemma 4.3. *Suppose $M \cup M_{\text{opt}}$ has a component that is an alternating path of length three, with the M-edge mw in the middle. Then $w' = M_{\text{opt}}(m)$ is matched in M'.*

Proof. Let $m' = M_{\text{opt}}(w)$ (see Figure 4) and suppose w' was single at the end of Phase 1 (i.e., w' is single in M'). As this is a component of $M \cup M_{\text{opt}}$, clearly both m' and w' are single in M, and moreover, as matched men never become single in Phase 2, m' is also single in M'.

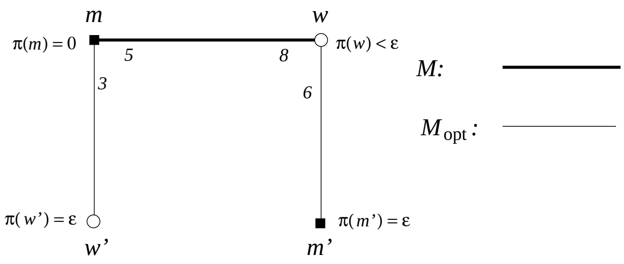

Figure 4: A path of length three in $M \cup M_{\text{opt}}$

First we observe that as w' is single in M', m did not propose to her during Phase 1, so $\pi(m) = 0$ (as $\pi(m)$ could only be positive after m searched over his strict list). However m' remained single, so $\pi(m') = \varepsilon$ at the end of the algorithm.

In Phase 2, w did not propose to m' (m' remained single, thus he did not receive any proposals), so $\pi(w) \leq \varepsilon/2$. Next we use $M(w) = m$, and we consider two cases. If $M'(w) = m$ then in Phase 1, when w rejected m' the last time, she had $\text{pri}'(w, m) \geq \text{pri}'(w, m') = \text{pri}(w, m') + \varepsilon$, so that in this case $\text{pri}(w, m) > \text{pri}(w, m')$. Otherwise, if $M'(w) \neq m$ then in Phase 2 w started to make proposals from the beginning of her strict list (that was made with respect to pri$'$ after Phase 1), but she did not propose to m', so $\text{pri}'(w, m) \geq \text{pri}'(w, m')$ also implying $\text{pri}(w, m) > \text{pri}(w, m')$.

At the beginning of Phase 2, $\pi(w')$ was set to ε, and w' remained single. This means that w' proposed to m and m rejected her. Consequently $\text{pri}'(m, w) \geq \text{pri}'(m, w')$, thus $\text{pri}(m, w) > \text{pri}(m, w')$. These arguments show that mw is blocking for M_{opt}, a contradiction. □

We continue the proof of the theorem. Let SM denote the set of single men at the end of the algorithm. First note, that men in SM were also single after Phase 1, since in Phase 2 men's positions do not decline. Let $\widehat{SM} \subseteq SM$ denote the set of those single men who are matched in M_{opt}. Observe that for each man $m \in \widehat{SM}$, woman $M_{\text{opt}}(m)$ exists and is matched in both M' and M (at the end of any Phase at least one person in any acceptable pair is matched). We further partition $\widehat{SM}$ as follows. Let SM_1 consist of each man $m \in \widehat{SM}$, for whom man $M(M_{\text{opt}}(m))$ is matched in M_{opt}; and $SM_2 := \widehat{SM} \setminus SM_1$. Let $SM_1^1 := \{m \in SM_1 : M_{\text{opt}}(M(M_{\text{opt}}(m))) \text{ is matched in } M\}$ and $SM_1^2 := SM_1 \setminus SM_1^1$. By Lemma 4.3, for every man m in SM_1^2, woman $M_{\text{opt}}(M(M_{\text{opt}}(m)))$ is matched in M' (i.e., at the end of Phase 1). The next lemma plays a crucial role in the proof of the theorem.

Lemma 4.4.

$$|SM_1| \leq \frac{2}{3} \cdot |M|$$

Proof. Case 1 $\quad |SM_1^1| \geq |SM_1|/2$.

We form clubs, every club is led by a man in SM_1 and has one or two other men who are matched in M. For every man $m \in SM_1$ the second member of his club is $M(M_{\text{opt}}(m))$. For each man $m \in SM_1^1$, his club contains a third member: $M(M_{\text{opt}}(M(M_{\text{opt}}(m))))$. We claim that these clubs are pairwise disjoint.

We formed one club for each man in SM_1 so it is enough to prove that any man m' who is matched in M belongs to at most one club. If $M(m')$ is single in M_{opt} then m' is not a member of any club. If $m = M_{\text{opt}}(M(m')) \in SM$, then either $m \in SM_1$ and m' belongs to m's club or otherwise m' has no club at all. In the other case ($m \notin SM$), m' belongs to the club of $m^* = M_{\text{opt}}(M(M_{\text{opt}}(M(m'))))$ as a third member if m^* exists and $m^* \in SM_1^1$; and m' has no club otherwise.

Let MM denote the set of men who are matched in M. We have

$$|M| = |MM| \geq |SM_1| + |SM_1^1| \geq \frac{3}{2} \cdot |SM_1|.$$

Case 2 $|SM_1^2| > |SM_1|/2$.

In this case we form different clubs, here the non-leader members will be men matched in M'. For every man $m \in SM_1$ the second member of his club is $M'(M_{\text{opt}}(m))$. For each man $m \in SM_1^2$, his club contains a third member: $M'(M_{\text{opt}}(M(M_{\text{opt}}(m))))$. We claim that these clubs are also pairwise disjoint.

If $M'(m')$ is single in M_{opt} then m' is not a member of any club. If $m = M_{\text{opt}}(M'(m')) \in SM$, then either $m \in SM_1$ and m' belongs to m's club or otherwise m' has no club at all. Otherwise, m' belongs to the club of $m^* = M_{\text{opt}}(M(M_{\text{opt}}(M'(m'))))$ as a third member if m^* exists and $m^* \in SM_1^2$; and m' has no club otherwise.

Let MM' denote the set of men who are matched in M'. As men matched after Phase 1 remain matched till the end, we have

$$|M| = |MM| \geq |MM'| \geq |SM_1| + |SM_1^2| \geq \frac{3}{2} \cdot |SM_1|.$$

□

We are ready to finish the proof of the theorem. Let MM_{opt} denote the set of men who are matched in M_{opt}. We claim that $|MM \cap MM_{\text{opt}}| \leq |MM| - |SM_2|$. This is true because $|SM_2|$ is the number of components of $M \cup M_{\text{opt}}$ isomorphic to a path with two edges and with a woman in the middle; and for each such path the M-matched man is single in M_{opt}.

$$|M_{\text{opt}}| = |MM_{\text{opt}}| = |MM \cap MM_{\text{opt}}| + |SM \cap MM_{\text{opt}}| \leq$$

$$\leq (|MM| - |SM_2|) + (|SM_1| + |SM_2|) \leq |M| + \frac{2}{3} \cdot |M| = \frac{5}{3} \cdot |M|.$$

□

5 Open Problems

Open Problem 1. Is it possible to improve the performance of GSA1 if we use smaller ε, increase extra scores more than once, and give extra scores to not only single men, but also to partners of each woman who is a neighbor of a single man?

Open Problem 2. Is it possible to improve the performance of GSA1 if we use the method of Irving and Manlove [5] *after* GSA1?

Open Problem 3. Is it possible to improve the performance of GSA2 if we use smaller ε, increase extra scores more than once, alternately for men and women? (For example with $\varepsilon < 1/N$ repeat the algorithm Ntimes, in the ith repetition increasing the extra scores of singles to $i\varepsilon$).

Open Problem 4. Is it possible to improve the performance of GSA2 if we use the method of Halldórsson et al. [3], or the method of Iwama, Miyazaki and Yamauchi [7] *after* GSA2?

5.1 Acknowledgement

I am grateful to Tamás Fleiner for his invaluable advice. I am also indebted to the referees of ESA'08.

References

[1] D. Gale, L. S. Shapley, College admissions and the stability of marriage *Amer. Math. Monthly* **69** (1962) pp. 9–15.

[2] M. M. Halldórsson, R. W. Irving, K. Iwama, D. F. Manlove, S. Miyazaki, Y. Morita, S. Scott, Approximability results for stable marriage problems with ties *Theor. Comput. Sci.* **306** (2003) pp. 431–447.

[3] M. M. Halldórsson, K. Iwama, S. Miyazaki, H. Yanagisawa, Improved approximation results for the stable marriage problem *ACM Trans. Algorithms* **3** No. 3, (2007) Article 30.

[4] M. M. Halldórsson, K. Iwama, S. Miyazaki, H. Yanagisawa, Randomized approximation of the stable marriage problem *Theor. Comput. Sci.* **325** (2004) pp. 439–465.

[5] R. W. Irving, D. F. Manlove, Approximation algorithms for hard variants of the stable marriage and hospitals/residents problems *Journal of Combinatorial Optimization* (2007) DOI: 10.1007/s10878-007-9133-x

[6] K. Iwama, D. F. Manlove, S. Miyazaki, Y. Morita, Stable marriage with incomplete lists and ties *Proceedings of the 26th International Colloquium on Automata, Languages and Programming* (1999) **LNCS 1644** pp. 443–452.

[7] K. Iwama, S. Miyazaki, N. Yamauchi, A 1.875-approximation algorithm for the stable marriage problem *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (2007) pp. 288–297.

[8] D. F. Manlove, R. W. Irving, K. Iwama, S. Miyazaki, Y. Morita, Hard variants of stable marriage *Theor. Comput. Sci.* **276** (2002) pp. 261–279.