Gale-Shapley Stable Marriage Problem Revisited: Strategic Issues and Applications
(Extended Abstract)

Chung-Piaw Teo[1]*, Jay Sethuraman[2]**, and Wee-Peng Tan[1]

[1] Department of Decision Sciences, Faculty of Business Administration, National University of Singapore
fbateocp@nus.edu.sg
[2] Operations Research Center, MIT, Cambridge, MA 02139
jayc@mit.edu

1 Introduction

This paper is motivated by a study of the mechanism used to assign primary school students in Singapore to secondary schools. The assignment process requires that primary school students submit a rank ordered list of six schools to the Ministry of Education. Students are then assigned to secondary schools based on their preferences, with priority going to those with the highest examination scores on the Primary School Leaving Examination (PSLE). The current matching mechanism is plagued by several problems, and a satisfactory resolution of these problems necessitates the use of a *stable matching* mechanism. In fact, the *student-optimal* and *school-optimal* matching mechanisms of Gale and Shapley [2] are natural candidates.

Stable matching problems were first studied by Gale and Shapley [2]. In a stable marriage problem we have two finite sets of players, conveniently called the set of men (M) and the set of women (W). We assume that every member of each set has strict preferences over the members of the opposite sex. In the *rejection* model, the preference list of a player is allowed to be incomplete in the sense that players have the option of declaring some of the members of the opposite sex as unacceptable; in the *Gale-Shapley* model we assume that preference lists of the players are *complete*. A matching is just a one-to one mapping between the two sexes; in the rejection model, we also include the possibility that a player may be unmatched, i.e. the player's assigned partner in the matching is himself/herself. The matchings of interest to us are those with the crucial *stability* property, defined as follows: A matching μ is said to be *unstable* if there is a man-woman pair, who both prefer each other to their (current) assigned partners in μ; this pair is said to block the matching μ, and is called a *blocking pair* for μ. A *stable matching* is a matching that is not unstable. The significance of stability is best highlighted by a system where acceptance of the proposed matching is

* Research supported by a NUS Research grant RP 3970021.
** Research supported by an IBM Fellowship.

G. Cornuéjols, R.E. Burkard, G.J. Woeginger (Eds.): IPCO'99, LNCS 1610, pp. 429–438, 1999.

voluntary. In such a setting, an unstable matching cannot be expected to remain intact, as the blocking pair(s) would soon discover that they could both improve their match by joint action; the man and the woman involved in a blocking pair could just “divorce” their respective partners and “elope.” To put this model in context, the students play the role of “women” and the secondary schools play the role of the “men.” Observe that there is a crucial difference between the stable marriage model as described here, and the problem faced by the primary school students in Singapore; in the latter, many “women” (students) can be assigned to the same “man” (secondary school), whereas in the former we require that the matching be one-to-one. In what follows, we shall restrict our attention to the one-to-one marriage model; nevertheless, the questions studied here, and the ideas involved in their resolution are relevant to the many-to-one marriage model as well.

One of the main difficulties with using the men-optimal stable matching mechanism is that it is manipulable by the women: some women can intentionally misrepresent their preferences to obtain a better stable partner. Such (strategic) questions have been studied for the stable marriage problem by mathematical economists and game theorists; essentially, this approach seeks to understand and quantify the potential gains of a deceitful player. Roth [7] proved that when the men-optimal stable-matching mechanism is used, a man will never be better off by falsifying his preferences, if all the other people in the problem reveal their true preferences. Falsifying preferences will at best result in the (original) match that he obtains when he reveals his true preferences. Gale and Sotomayor [4] showed that when the man-propose algorithm is used, a woman w can still force the algorithm to match her to her women-optimal partner, denoted by $\mu(w)$, by falsifying her preference list. The optimal cheating strategy for woman w is to declare men who rank below $\mu(w)$ as unacceptable. Indeed, the cheating strategy proposed by Gale and Sotomayor is optimal in the sense that it enables the women to obtain their women-optimal stable partner even when the man-propose mechanism is being adopted. Prior to this study, we know of no analogous results when the women are required to submit *complete* preference lists. This question is especially relevant to the Singapore school-admissions problem: All assignments are done via the centralized posting exercise, and no student is allowed to approach the schools privately for admission purposes. In fact, in the current system, students not assigned to a school on their list are assigned to schools according to some pre-determined criterion set by the Ministry. Thus effectively this is a matching system where the students are not allowed to remain “single” at the end of the posting exercise. To understand whether the stable matching mechanism is a viable alternative, we first need to know whether there is any incentive for the students to falsify their preferences, so that they can increase their chances of being assigned to “better” schools. The one-to-one version of this problem is exactly the question studied in this paper: in the stable marriage model with complete preferences, with the men-optimal matching mechanism, is there an incentive for the women to cheat? If so, what is the optimal cheating strategy for a woman? To our knowledge, the only result known about this problem is an example due to Josh Benaloh (cf. Gusfield and Irving [5]), in which the women lie by permuting their preference lists, and still manage to force the men-optimal matching mechanism to return the women-optimal solution.

2 Optimal Cheating in the Stable Marriage Problem

Before we derive the optimal cheating strategy we consider a (simpler) question: Suppose woman w is allowed to reject proposals. Is it possible for woman w to identify her women-optimal partner by observing the sequence of proposals in the men-propose algorithm? Somewhat surprisingly, the answer is yes! Our algorithm to compute the optimal cheating strategy is motivated by the observation that if a woman simply rejects all the proposals made to her, then the best (according to her true preference list) man among those who have proposed to her is her women-optimal partner. Hence by rejecting all her proposals, a woman can extract information about her best possible partner. Our algorithm for the optimal cheating strategy builds on this insight: the deceitful woman rejects as many proposals as possible, while remaining engaged to some man who proposed earlier in the algorithm. Using a backtracking scheme, the deceitful woman can use the matching mechanism repeatedly to find her optimal cheating strategy.

2.1 Finding Your Optimal Partner

We first describe algorithm *OP*—an algorithm to compute the women-optimal partner for w using the man-propose mechanism. (Recall that we do this under the assumption that woman w is allowed to remain single.)

Algorithm *OP*

1. Run the man-propose algorithm, and reject all proposals made to w. At the end, w and a man, say m, will remain single.
2. Among all the men who proposed to w in Step 1, let the best man (according to w) be m_1.

Theorem 1. *m_1 is the women-optimal partner for w.*

Proof: Let $\mu(w)$ denote the women-optimal partner for w. We modify w's preference list by inserting the option to remain single in the list, *immediately after* $\mu(w)$. (We declare all men that are inferior to $\mu(w)$ as unacceptable to w.) Consequently, in the man-propose algorithm, all proposals inferior to $\mu(w)$ will be rejected. Nevertheless, since there exists a stable matching with w matched to $\mu(w)$, our modification does not destroy this solution. It is also well known that the set of people who are single is the same for all stable matchings (cf. Roth and Sotomayor [8], pg. 42). Thus, w must be matched in *all* stable matchings with the modified preference list. The men-optimal matching for this modified preference list must match w to $\mu(w)$. In particular, $\mu(w)$ must have proposed to

w during the execution of the man-propose algorithm. Note that until $\mu(w)$ proposes to w, the man-propose algorithm for the modified list runs exactly in the same way as in Step 1 of OP. The difference is that Step 1 of OP will reject the proposal from $\mu(w)$, while the man-propose algorithm for the modified list will accept the proposal from $\mu(w)$, as w prefers $\mu(w)$ to being single. Hence, clearly $\mu(w)$ is among those who proposed to w in Step 1 of OP, and so $m_1 \geq_w \mu(w)$.

Suppose $m_1 >_w \mu(w)$. Consider the modified list in which we place the option of remaining single *immediately after* m_1. We run the man-propose algorithm with this modified list. Again, until m_1 proposes to w, the algorithm runs exactly the same as in Step 1 of OP, after which the algorithm returns a stable partner for w who is at least as good as m_1. This gives rise to a contradiction as we assumed $\mu(w)$ to be the best stable partner for w. ■

Observe that under this approach, the true preference list of w is only used to compare the men who have proposed to w. We do not need to know her exact preference list; we only need to know which man is the best among a given set of men, according to w. Hence the information set needed here to find the women-optimal partner of w is much less than that needed when the woman-propose algorithm is used. This is useful for the construction of the cheating strategy as the information on the "optimal" preference list is not given a-priori and is to be determined.

2.2 Cheating Your Way to a Better Marriage

Observe that the preceding procedure only works when woman w is allowed to remain single throughout the matching process, so that she can reject any proposal made to her in the algorithm. Suppose we do not give the woman an option to declare any man as unacceptable. How do we determine her best stable partner? This is essentially a restatement of our original question: who is the best stable partner woman w can have when the man-propose algorithm is used and when she can lie only by permuting her preference list.

A natural extension of Algorithm OP is for woman w to: (i) accept a proposal first, and then reject all future proposals. (ii) From the list of men who proposed to w but were rejected, find her most preferred partner; repeat the Gale-Shapley algorithm until the stage when this man proposes to her. (iii) Reverse the earlier decision and accept the proposal from this most preferred partner, and continue the Gale-Shapley algorithm by rejecting all future proposals. (iv) Repeat (ii) and (iii) until the woman cannot find a better partner from all other proposals. Unfortunately, this elegant strategy does not always yield the best stable partner a woman can achieve under our model. The reason is that this greedy improvement technique does not allow for the possibility of rejecting the current best partner, in the hope that this rejection will trigger a proposal from a *better* would-be partner. Our algorithm in this paper does precisely that. Let $P(w) = \{m_1, m_2, \ldots m_n\}$ be the true preference list of woman w, and let $P(m, w)$ be a preference list for w that returns m as her men-optimal partner. Our algorithm constructs $P(m, w)$ iteratively, and consists of the following steps:

1. Run the man-propose algorithm with the true preference list $P(w)$ for woman w. Keep track of all men who propose to w. Let the men-optimal partner for w be m, and let $P(m, w)$ be the true preference list $P(w)$.
2. Suppose m_j proposed to w in the Gale-Shapley algorithm. By moving m_j to the front of the list $P(m, w)$, we obtain a preference list for w such that the men-optimal partner will be m_j. Let $P(m_j, w)$ be this altered list. We say that m_j is a *potential* partner for w.
3. Repeat step 2 for every man who proposed to woman w in the algorithm; after this, we say that we have *exhausted* man m, the men-optimal partner obtained with the preference list $P(m, w)$.
4. If a potential partner for w, say man u, has not been exhausted, run the Gale-Shapley algorithm with $P(u, w)$ as the preference list of w. Identify new possible partners and their associated preference lists, and classify man u as exhausted.
5. Repeat Step 4 until all possible partners of w are exhausted. Let N denote the set of all possible (and hence exhausted) partners for w.
6. Among the men in N let m_a be the man woman w prefers most. Then $P(m_a, w)$ is an optimal cheating strategy for w.

The men in the set N at the end of the algorithm have the following crucial properties:

– For each man m in N, there is an associated preference list for w such the Gale-Shapley algorithm returns m as the men-optimal partner for w with this list.
– All other proposals in the course of the Gale-Shapley algorithm come from other men in N. (Otherwise, there will be some possible partners who are not exhausted.)

With each run of the Gale-Shapley algorithm, we exhaust a possible partner, and so we need at most n Gale-Shapley algorithms before termination.

Theorem 2. $\pi = P(m_a, w)$ *is an optimal strategy for woman* w.

Proof: (by contradiction) We use the convention that $r(m) = k$ if man m is the k^{th} man on woman w's list. Let $\pi^* = \{m_{p1}, m_{p2}, \ldots, m_{pn}\}$ be the preference list that gives rise to the *best* stable partner for w under the man-propose algorithm. Let this man be denoted by m_{pb}, and let woman w strictly prefer m_{pb} to m_a (under her true preference list). Recall that we use $r(m)$ to represent the rank of m under the true preferences of w; by our assumption, $r(m_{pb}) < r(m_a)$, i.e., m_{pb} is ranked higher than m_a. Observe that the order of the men who do not propose to woman w is irrelevant and does not affect the outcome of the Gale-Shapley's algorithm. Furthermore, men of rank higher than $r(m_{pb})$ do not get to propose to w, otherwise we can cheat further and improve on the best partner for w, contradicting the optimality of π^*. Thus we can arbitrarily alter the order of these men, without affecting the outcome. Without loss of generality, we may assume that $1 = r(m_{p1}) < 2 = r(m_{p2}) < \ldots < q = r(m_{pb})$.

Since $r(m_{pb}) < r(m_a)$, m_a will appear anywhere after m_{pb} in π^*: thus, m_a can appear in any position from m_{pb+1} to m_{pn}.

Now, we modify π^* such that all men who (numerically) rank lower than m_a but higher than m_{pb} (under true preferences) are put in order according to their ranks. This is accomplished by moving all these men before m_a in π^*. With that alteration, we obtain a new list $\tilde{\pi} = \{m_{q1}, m_{q2}, \ldots, m_{qn}\}$ such that:

(i) $1 = r(m_{q1}) < 2 = r(m_{q2}) < \ldots < s = r(m_{qs})$.

(ii) $m_{q1} = m_{p1} \ldots m_{qb} = m_{pb}$, where the position of those men who rank higher than m_{pb} is unchanged.

(iii) $r(m_a) = s + 1$, $m_a \in \{m_{qs+1}, m_{qs+2}, \ldots m_{qn}\}$.

(iv) The men in the set $\{m_{qs+1}, m_{qs+2}, \ldots m_{qn}\}$ retain their relative position with respect to one another under π^*.

Note that the men-optimal partner of w under $\tilde{\pi}$ cannot come from the set $\{m_{qs+1}, m_{qs+2}, \ldots m_{qn}\}$. Otherwise, since the set of men who proposed in the course of the algorithm must come from $\{m_{qs+1}, m_{qs+2}, \ldots m_{qn}\}$, and since the preference list π^* retains the relative order of the men in this set, the same partner would be obtained under π^*. This leads to a contradiction as π^* is supposed to return a better partner for w. Hence, we can see that under $\tilde{\pi}$, we already get a better partner than under π.

Now, since the preference list π returns m_a with $r(m_a) = s + 1$, we may conclude that the set N (obtained from the final stage of the algorithm) does not contain any man of rank smaller than $s+1$. Thus $N \subseteq \{m_{qs+1}, m_{qs+2}, \ldots m_{qn}\}$. Suppose $m_{qs+1}, m_{qs+2}, \ldots, m_{qw}$ do not belong to the set N, and m_{qw+1} is the first man after m_{qs} who belongs to the set N. By construction of N, there exists a permutation $\hat{\pi}$ with m_{qw+1} as the stable partner for w under the men-optimal matching mechanism. Furthermore, all of those who propose to w in the course of the algorithm are in N, and hence they are no better than m_a to w. Furthermore, all proposals come from men in $\{m_{qw+1}, m_{qw+2}, \ldots m_{qn}\}$, since $N \subseteq \{m_{qs+1}, m_{qs+2}, \ldots m_{qn}\}$.

By altering the order of those who did not propose to w, we may assume that $\hat{\pi}$ is of the form $\{m_{q1}, m_{q2}, \ldots, m_{qs-1}, m_{qs}, \ldots, m_{qw}, m_{qw+1}, \ldots\}$, where the first $qw+1$ men in the list are identical to those in $\tilde{\pi}$. But, the men-optimal stable solution obtained using $\hat{\pi}$ must also be stable under $\tilde{\pi}$, since w is match to m_{qw+1}, and the set of men she strictly prefers to m_{qw+1} is identical in both $\hat{\pi}$ and $\tilde{\pi}$. This is a contradiction as $\tilde{\pi}$ is supposed to return a men-optimal solution better than m_a. Thus π^* does not exist, and so π is optimum and m_a is the best stable partner w can get by permuting her preference list. ■

We now present an example to illustrate how our heuristic works.

Example 1: Consider the following stable marriage problem:

$$\begin{array}{c|ccccc} 1 & 2&3&4&5&1 \\ 2 & 3&4&5&1&2 \\ 3 & 5&1&4&2&3 \\ 4 & 3&1&2&4&5 \\ 5 & 1&5&2&3&4 \end{array} \qquad \begin{array}{c|ccccc} 1 & 1&2&3&5&4 \\ 2 & 2&1&4&5&3 \\ 3 & 3&2&5&1&4 \\ 4 & 4&5&1&2&3 \\ 5 & 5&1&2&3&4 \end{array}$$

True Preferences of the Men True Preferences of the Women

We construct the optimal cheating strategy for woman 1.

– Step 1: Run Gale-Shapley with the true preference list for woman 1; her men-optimal partner is man 5. Man 4 is the only other man who proposes to her during the Gale-Shapley algorithm. So $P(\text{man5}, \text{woman1}) = (1, 2, 3, 5, 4)$.
– Step 2-3: Man 4 is moved to the head of woman 1's preference list; i.e., $P(\text{man4}, \text{woman1}) = (4, 1, 2, 3, 5)$. Man 5 is exhausted, and man 4 is a potential partner.
– Step 4: As man 4 is not yet exhausted, we run the Gale-Shapley algorithm with $P(\text{man4}, \text{woman1})$ as the preference list for woman 1. Man 4 will be exhausted after this, and man 3 is identified as a new possible partner, with $P(\text{man3}, \text{woman1}) = (3, 4, 1, 2, 5)$.
– Repeat Step 4: As man 3 is not yet exhausted, we run Gale-Shapley with $P(\text{man3}, \text{woman1})$ as the preference list for woman 1. Man 3 will be exhausted after this. No new possible partner is found, and so the algorithm terminates.

Example 1 shows that woman 1 could cheat and get a partner better than the men-optimal solution. However, her women-optimal partner in this case turns out to be man 1. Hence Example 1 also shows that woman 1 cannot always assure herself of getting the women-optimal partner through cheating, in contrast to the case when rejection is allowed in the cheating strategy.

3 Strategic Issues in the Gale-Shapley Problem

By requiring the women to submit complete preference lists, we are clearly restricting their strategic options, and thus many of the strong structural results known for the model with rejection may not hold in this model. This is good news, for it reduces the incentive for a woman to cheat. In the rest of this section, we present some examples to show that the strategic behaviour of the women can be very different under the models with and without rejection.

3.1 The Best Possible Partners (Obtained from Cheating) May Not Be Women-Optimal

In the two-sided matching model with rejection, it is not difficult to see that the women can always force the man-propose algorithm to return the women-optimal solution (e.g. each woman rejects all those who are inferior to her women-optimal partner). In our model, where rejection is forbidden, the influence of the women is far less, even if they collude. A simple example is when each woman is ranked first by exactly one man. In this case, there is no conflict among the men, and in the men-optimal solution, each man is matched to the woman he ranks first. (This situation arises whenever each man ranks his men-optimal partner as his first choice.) In this case, the algorithm will terminate with the men-optimal solution, regardless of how the women rank the men in their lists. So ruling out the strategic option of remaining single for the women significantly affects their ability to change the outcome of the game by cheating.

By repeating the above analysis for all the other women in Example 1, we conclude that the best possible partner for woman 1, 2, 3, 4, and 5 are respectively man 3, 1, 2, 4, and 3. An interesting observation is that woman 5 cannot benefit by cheating alone (she can only get her men-optimal partner no matter how she cheats). However, if woman 1 cheats using the preference list $(3, 4, 1, 2, 5)$, woman 5 will also benefit by being matched to man 5, who is first in her list.

3.2 Multiple Strategic Equilibria

Suppose each woman w announces a preference list $P(w)$. The set of strategies $\{\pi(1), \pi(2), \ldots, \pi(n)\}$ is said to be in strategic equilibrium if none of the women has an incentive to deviate unilaterally from this announced strategy. It is easy to see that if a woman benefits from announcing a different permutation list (instead of her true preference list), then every other woman would also benefit, i.e. every other woman will get a partner who is at least as good as her men-optimal partner (cf. Roth and Sotomayor [8]).

Theorem 3. *If a single woman can benefit by cheating, then the game has multiple strategic equilibria.*

Proof: A strategic equilibrium can be constructed by repeating the proposed cheating algorithm iteratively, improving the partner for some woman at each iteration. (Notice that the partner of a woman at the end of iteration j is at least as good as her partner at the beginning of the iteration j.) The algorithm will thus terminate at a strategic equilibrium, where at least one woman will be matched to someone whom she (strictly) prefers to her men-optimal partner. Another strategic equilibrium is obtained if each woman w announces a list of the form $\{m_1, m_2, \ldots, m_n\}$, with m_1 being her men-optimal partner and $m_2, m_3, \ldots, m_n$ in the same (relative) order as in her true preference list. Clearly the man-propose algorithm will match woman w to m_1, since moving m_1 to the front of w's preference list does not affect the sequence of proposals in the man-propose algorithm. No woman can benefit from cheating, as all other women are already matched to their announced first-ranked partner. Thus we have constructed two strategic equilibria. ■

3.3 Does It Pay To Cheat?

Roth [7] shows that under the man-propose mechanism, the men have no incentives to alter their true preference lists. In the rejection model, however, Gale and Sotomayor [3] show that a woman has an incentive to cheat as long as she has at least two distinct stable partners. Pittel [6] shows that the average number of stable solutions is asymptotic to $nlog(n)/e$, and with high probability, the rank of the women-optimal and men-optimal partner for the woman are respectively $log(n)$ and $n/log(n)$. Thus in typical instances of the stable marriage game under the rejection model, most of the women will not reveal their true preference lists.

Many researchers have argued that the troubling implications from these studies are not relevant in practical stable marriage game, as the model assumes that the women have full knowledge of each individual's preference list and the set of all the players in the game. For the model we consider, it is natural to ask whether it pays (as in the rejection model) for a typical woman to solicit information about the preferences of all other participants in the game. We run the cheating algorithm on 1000 instances, generated uniformly at random, for $n = 8$ and the number of women who benefited from cheating is tabulated in Table 1.

Number of Women who benefited	0	1	2	3	4	5	6	7	8
Number of observations	740	151	82	19	7	1	0	0	0

Table 1

Interestingly, the number of women who can gain from cheating is surprisingly low. In fact, in 74% of the instances, the men-optimal solution is their only option, no matter how they cheat. The average percentage of women who benefit from cheating is merely 5.06%.

To look at the typical strategic behaviour on larger instances of the stable marriage problem, we run the heuristic on 1000 random instances for $n = 100$. The cumulative plot is shown in Figure 1. In particular, in more than 60% of the instances at most 10 women (out of 100) benefited from cheating, and in more than 96% of the instances at most 20 women benefited from cheating. The average number of women who benefited from cheating is 9.52%. Thus, the chances that a typical woman can benefit from acquiring complete information (i.e., knowing the preferences of the other players) is pretty slim in our model.

We have repeated the above experiment for large instances of the Gale-Shapley model. Due to computational requirements, we can only run the experiment on 100 random instances of the problem with 500 men and women. Again the insights obtained from the 100 by 100 cases carry over: the number of women who benefited from cheating is again not more than 10% of the total number of the women involved. In fact, the average was close to 6% of the women population in the problem. This suggests that the number of women who can benefit from cheating in the Gale-Shapley model with n women grows at a rate which is slower than a linear function of n. However, detailed probabilistic

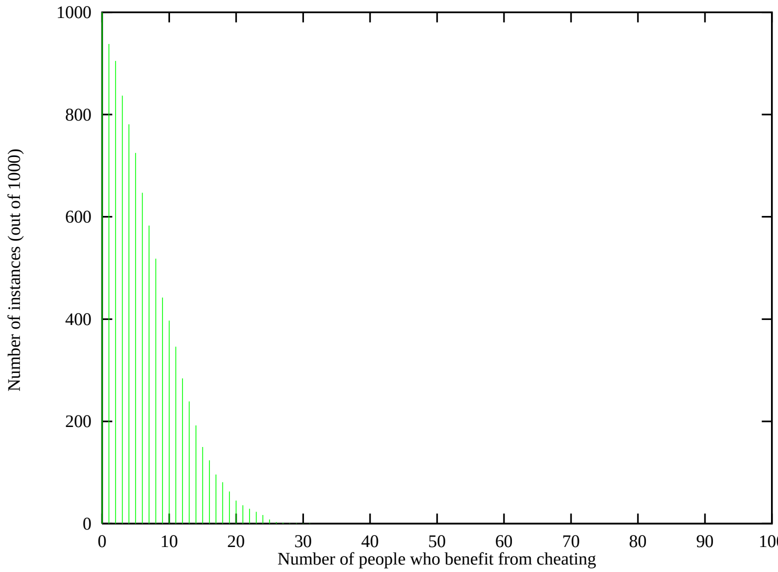

Fig. 1. Benefits of cheating: cumulative plot for $n = 100$

analysis of this phenomenon is a challenging problem that is beyond the scope of the present paper.

A practical advice for all women in the stable marriage game, using men-optimal matching mechanism: “don’t bother to cheat !”

References

1. Dubins, L.E. and D.A. Freedman (1981), “Machiavelli and the Gale- Shapley Algorithm.“ American Mathematical Monthly, 88, pp485-494.
2. Gale, D., L. S. Shapley (1962). College Admissions and the Stability of Marriage. American Mathematical Monthly 69 9-15.
3. Gale, D., M. Sotomayor (1985a). Some Remarks on the Stable Matching Problem. Discrete Applied Mathematics 11 223-232.
4. Gale, D., M. Sotomayor (1985b). Ms Machiavelli and the Stable Matching Problem. American Mathematical Monthly 92 261-268.
5. Gusfield, D., R. W. Irving (1989). The Stable Marriage Problem: Structure and Algorithms, MIT Press, Massachusetts.
6. Pittel B. (1989), the Average Number of Stable Matchings, SIAM Journal of Discrete Mathematics, 530-549.
7. Roth, A.E. (1982). The Economics of Matching: Stability and Incentives. Mathematics of Operations Research 7 617-628.
8. Roth, A.E., M. Sotomayor (1991). Two-sided Matching: A Study in Game-theoretic Modelling and Analysis, Cambridge University Press, Cambridge.