Algorithmic Game Theory, Summer 2018 — Lecture 24 (5 pages)

Stable Matching

Instructor: Thomas Kesselheim

Continuing our study of mechanism design without money, we will turn to *stable matching* today. The classical motivation comes from (heterosexual, monogamous) marriages. There is a set of *men* U and a set of *women* V, which have to be matched to each other. That is we have to find a matching M, which assigns each man and each woman at most one partner.

Every man has a *preference order* over the women and every woman has a preference order over the men. For example, we could have $U = \{x, y, z\}$, $V = \{a, b, c\}$. Now the preference lists could be

$$x : a > b > c, \quad y : b > c > a, \quad z : a > c > b,$$
$$a : y > z > x, \quad b : y > z > x, \quad c : x > y > z\,.$$

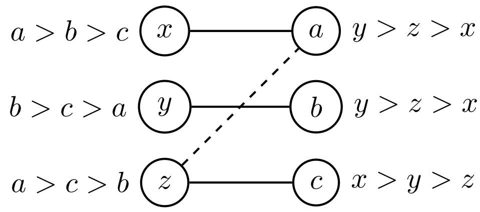
These preferences give a complete order among the persons on the other side. They are arbitrary and don't have to be consistent with the preferences of the others in any way. So, in this case, x ideally would like to be matched to a and rather b than c. However, in any case *being matched is preferred over not being matched.*

Definition 24.1. *Given a matching M, a pair of a man and a woman (u, v) is a* blocking pair *if u prefers v over the current matching partner and v prefers u over the current matching partner.*

That is, if (u, v) is a blocking pair, u and v prefer being matched to each other compared to their current match. If, for example, both u and v are unmatched, then (u, v) is automatically a blocking pair because being unmatched is the least desirable situation. Overall, there are four kinds of blocking pairs.

Definition 24.2. *We call a matching M* stable *if there is no blocking pair (u, v).*

Example 24.3. *Consider the above preferences and the following matching drawn in solid lines.*



Now the dashed edge is a blocking pair. This means, the matching is not stable.

Note that in the definition the roles of men and women are interchangeable; we might as well swap their roles in what follows. Besides, the motivation by marriages is maybe somewhat weak. A better one is the assignment of job applicants to jobs. Each potential employer and employee has a preference over the respective other side.

1 The Gale-Shapley Algorithm

Our first result is that there is always a stable matching. Our proof will be constructive: We will give an algorithm that computes one.

Men-proposing algorithm

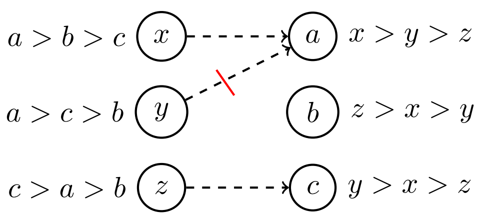
- Start with an empty matching M.
- Repeat
  - Each man proposes to his most preferred woman who has not rejected him yet (if any).
  - Each woman is tentatively matched to her favorite proposer (if any) and rejects the rest.

  until no rejections occur.
- Set M to be the tentative matches.

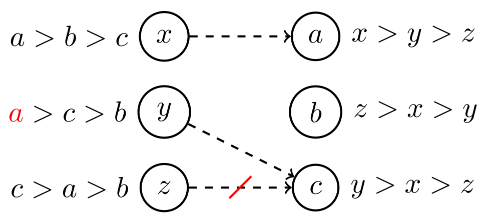
Let us have a look at how the men-proposing algorithm works by inspecting the following example.

Example 24.4. *Consider a set of men $U = \{x, y, z\}$ and a set of women $V = \{a, b, c\}$. Their respective preference orders are stated below. Each of the four diagrams illustrates a single iteration of the men-proposing algorithm. The dashed lines indicate proposals by men and the rejections made by the women are depicted by the red lines. Since there are no further rejections in the fourth iteration, the green edges represent the final matching M.*

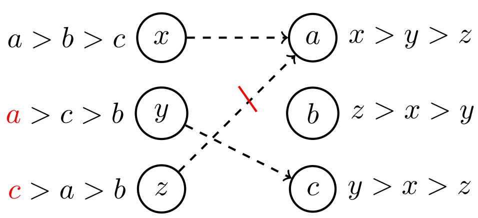
First iteration:



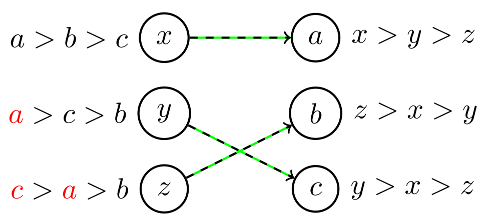
Second iteration:



Third iteration:



Fourth iteration:



Now, we will prove that the men-proposing algorithm yields a stable matching for arbitrary preference profiles.

Theorem 24.5. *The men-proposing algorithm always terminates with a stable matching after at most $|U| \cdot |V| + 1$ iterations.*

Proof. Let us first observe that the algorithm indeed terminates after $|U| \cdot |V| + 1$ iterations. This is simply due to the fact that in every iteration except the last one at least one woman rejects one man. This man will never propose to the woman again, so this exact rejection will never occur again. Before we show that the outcome of the algorithm is stable, let us record the following observations:

1. If a woman is tentatively matched to a man in one iteration, this same man will propose to her again in the next iteration. This is because she does not reject him (they are tentatively matched) and every more preferred woman will still have rejected him in the next iteration.

2. As a consequence, every woman that is tentatively matched at some point will remain tentatively matched for the rest of the execution. This does not mean that the match cannot change. However, a woman's tentative match can only improve from iteration to iteration.

Now, we are ready to show that there will be no blocking pair (u, v) in the final matching M. If u prefers v to his match in M or is not matched at all, u has to have proposed to v at some point. This means that v rejected u and is matched to some other man that she prefers because of the second observation. Therefore (u, v) is not a blocking pair. □

2 Who gets matched?

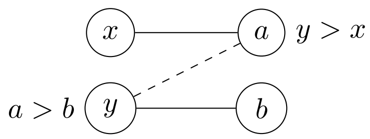
It is easy to see that often there is more than one stable matching. As there is no objective function, no one of them is clearly better than the other. As we will show, the one computed by the men-proposing algorithm is the most favorable in one sense and the least favorable in another sense. The first statement is that men always get matched with their "best choices".

Consider any agent i. We call an agent j from the other side *attainable* for i if there is a stable matching in which i gets matched to j.

Theorem 24.6. *In the outcome of the men-proposing algorithm, every man $u \in U$ gets matched to his most preferred attainable woman.*

Proof. Suppose that the theorem statement is false. Then, at some point, a man x gets rejected by his most preferred attainable woman, who we call a. This has to be because some man y also proposed to a and a prefers y over x. Because x was the first man to be rejected from his most preferred attainable woman, y likes a as least as much as his own most preferred attainable woman.

As a is attainable for x, there has to be a stable matching in which x is matched to a and y is matched to some other woman b.

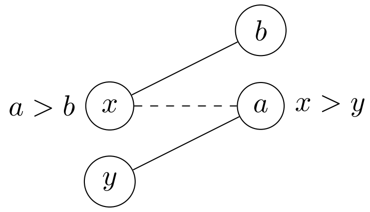


We now show that (y, a) is a blocking pair. We already argued that a prefers y over x. But as b is attainable for y, he also likes a better than b. This is a contradiction. □

For women, the situation is exactly opposite. They like the outcome least among all possible stable matchings.

Theorem 24.7. *In the outcome of the men-proposing algorithm, every woman $v \in V$ gets matched to her least preferred attainable man.*

Proof. Suppose for a contradiction that some woman a gets matched to a man x who she prefers to her least preferred attainable man y. Then, there has to be a stable matching in which y gets matched to a and x gets matched to $b \neq a$ (or not matched at all).



By Theorem 24.6, x likes a better than b. By our definition, a likes x better than y. So, (x, a) is a blocking pair for this matching, which is a contradiction. □

Corollary 24.8. *If a woman $v \in V$ is assigned to the same man in both the men-proposing and the women-proposing version of the algorithm, then this is the only attainable man for her.*

A consequence of the mentioned "men optimality" is that in any stable matching always the same men and women get a matching partner.

Theorem 24.9. *In any stable matching, the set women (and men) who are matched to the other side is the same.*

Proof. We show this statement for the women only. To show it for the men, switch their roles. For the definition of stability (unlike the algorithm) this does not change anything.

Observe that every stable matching always includes exactly $\min\{|U|, |V|\}$ edges: If on both sides there is an unmatched agent, they are a blocking pair.

To prove the theorem, consider a woman a who gets matched to x by the men-proposing algorithm. Suppose there is another stable matching M' in which a is not matched. If x is not matched either, they are already a blocking pair. If x is matched to some other woman $b \neq a$, we know that he prefers a over b because a is his most preferred attainable woman. Therefore, (x, a) is also a blocking pair in this case. □

3 Incentive Compatibility

We now come back to our usual question: Can agents benefit from lying? In this case, this means: Can agents improve their matches from reporting preferences other than their actual ones? For the men-proposing algorithm, the answer is both yes and no. On the one hand, men can never benefit from misreporting while women sometimes can.

Theorem 24.10. *Regardless of the other agents' preferences, in the men-proposing algorithm, no man can increase his match by misreporting his preference list.*

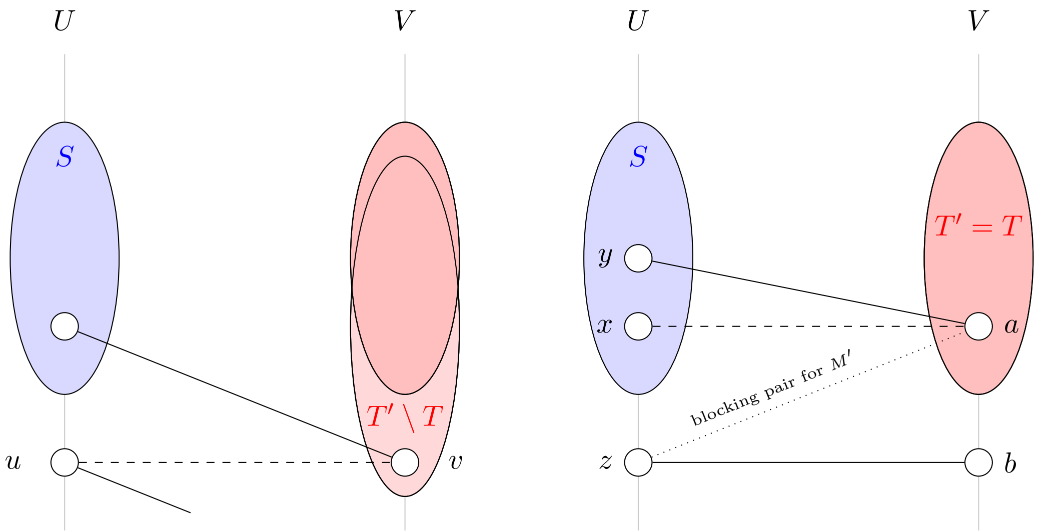
We show this theorem by proving the following lemma, which is an even stronger statement.

Lemma 24.11. *Given fixed preferences, let M be the outcome of the men-proposing algorithm, and let M' be any other matching. Let S be the set of men who prefer their match in M' to their match in M. If $S \neq \emptyset$, there is a man $u \notin S$ and a woman v such that (u, v) is a blocking pair with respect to M'.*

Proof. Let T be the set of women that S is matched to in M, T' the respective set in M'.

We consider two cases. The first case is $T \neq T'$. Note that it is impossible that T' is a strict subset of T. Suppose to the contrary that $T' \subset T$. Then, there has to be a man of S who does not get matched in M' but gets matched in M. But, from this we can infer a contradiction to the assumption that each man of S prefers his match in M' to his match in M. Therefore, there has to be some $v \in T' \setminus T$. Clearly, v also has to be matched in M, otherwise M could not be stable. Let $u \notin S$ be v's partner in M. We claim that (u, v) is a blocking pair with respect to M'. As $u \notin S$, u prefers v to his match in M'. Furthermore, v's partner in M' must also have proposed to v when running the algorithm but was rejected. That is, v prefers u over this other partner.

The second case is $T = T'$. When running the algorithm to get M, every woman in T gets proposals from S, one of which will lead to the eventual match. Let a be the woman who receives the last such proposal. Let x be her match in M. Hence, this last proposal is made by x. Furthermore, let y be her match in M'. Observe that y must have proposed to her before because he prefers a over his match in M. Therefore, a is tentatively matched when x proposes.



(a) Case $T' \neq T$ (b) Case $T' = T$

Figure 1: Matches of M are depicted by dashed lines and matches of M' by solid lines.

However, this tentative match has to be some man $z \notin S$. If $z \in S$, then there would be another proposal between S and T, contradicting the assumption that this is the last such proposal. We claim that (z, a) is blocking with respect to M'. We know that z prefers a to his final match in M. As $z \notin S$, this also hold for his match in M'. Furthermore, recall that y (a's match in M') also proposed to a but was rejected in favor of z. So, also a prefers z over her match in M'. □

Now, let us utilize the last lemma in order to prove Theorem 24.10.

Proof. Given fixed preferences of men and women, let M be the outcome of the men-proposing algorithm. Moreover, suppose that a nonempty set S_0 of men misrepresent their preferences. Claim: There is no stable matching M' for the resulting preference profile where all men in S_0 obtain strictly better matches than in M. Suppose to the contrary that there exists such a stable matching M'. Then, S as defined in Lemma 24.11 contains S_0. Hence, there exists a pair (u, v) that is blocking for both preference profiles because $u \notin S$. □

4 Further Directions

There are some interesting further questions related to stable matching. For example, in the *hospital-residents problem*, one side of the graph may get multiple matches (up to some capacity). These vertices represent hospitals, which doctors are assigned to. Indeed in the US, there is a system to assign doctors in such a way and the allocation using matching algorithms dates back to even the time before computers.

Another problem is the *stable roommates* problem. Now, the graph is not bipartite. Each agent has a preference list over all vertices in this graph and again the goal is to find a matching. The notion of stability extends naturally by the absence of a blocking pair. Unfortunately, there is not always a stable matching in this case.

Further Reading

- Section 10.1 - 10.3 in the Karlin/Peres book
- Section 10.4 in the AGT book
- Tim Roughgarden's lecture notes `http://theory.stanford.edu/~tim/f13/l/l10.pdf` and lecture video `https://youtu.be/NT07sILhsv4?t=56m54s`