The final problem that we will study during this last lecture is the stable matching problem viewed from a game theoretic standpoint. The stable matching problem arises in many practical settings the most prominent of which is the assignment of hospital residencies. Each candidate knows his preferred hospitals and each hospital knows the candidates it prefers. The goal is to find a "good" assignment of candidates to hospitals, where "good" will be defined shortly. We will start by overviewing the problem and the well known proof of existence of stable matchings. Then we will shift to more game theoretic aspects of the problem.

1 The Stable Matching Problem

In this problem we are given n men and n women and each person has a full perference list of the people of opposite gender. Our goal is to find a pairing (matching) of men and women that in some sense is stable. More formally we say a matching is *unstable* if there exists a pair of couples $(m_1, w_1), (m_2, w_2)$ such that m_1 and w_2 perfer each other to their current partners. A matching is stable if there are no unstable couples. Our goal once again is to find stable matchings.

There is a well known algorithm for finding stable matchings. The algorithm consists of men proposing women in order of preference and women holding on to their best proposal so far (obviously the symmetric algorithm of women proposing also works). It is well known that this algorithm finds stable matchings and the the final matching doesn't depend on the order of proposals (notice that the algorithm above was underspecified in that regard). Also all men are paired with their best stable partner in the outcome of this algorithm and all women are paired with their worst stable partner.

2 The Game Theory Question

Is this mechanism truthful? Is it truthful for men? Is it truthful for women? In the rest of this lecture we will address these questions in a couple of different settings.

First, it can be shown (though it is nontrivial) that the above mechanism is truthful for men. However, it is not truthful for women. To anaylize the truthfulness of the above mechanism with respect to women we will consider a more restricted setting where men have a partial preference list of at most k women and if none of those women prefer them then they'd rather stay single. Women on the other hand do have the full preferece list of men. The motivation of this is that men often indeed can't rank all the women in the world, yet women never need to have a preference list - all they need to be able to do is to decide between two alternatives, which they surely can.

Again it can be shown that the mechanism produces a stable matching, where the notion of instability has to be modified to include running off with a single person. Similarly it can be shown that again the order of proposals is insignificant, and that men are paired with their best partners in any stable matching and women are paired with their worst partners (again we have to adjust best and worst to account for single people).

For this model as we will see with a small example the algorithm above is not truthful for women (it can be shown to be truthful for men). There is in fact a theorem stating that there is no truthful mechanism for this game.

The small example is the following - there are three men m_1, m_2, m_3 and three women w_1, w_2, w_3. and the preference list for w_2 is (m_3, m_1, m_2). Let us assume that the current couples are (m_i, w_i), w_2 is currently being offered by m_1, and w_3 prefers m_1 to all other men. If w_2 forgoes the offer by m_1 and stays with m_2 then m_1 will offer to w_3 which will open m_3 who will offer to w_2 thus w_2 will get his best choice. Therefore the mechanism isn't truthful.

But not all hope is lost. We can still show some positive things. In parcitular we will consider the case when all men choose $\leq k$ women in their preference lists and do so randomly. We will show (outline the proof) that in this scenario the number of women who would benefit from lying depends only on k and not on the number of total people.

Theorem 1 *Consider the problem instance where we are given a full preference lists of women and men choose at most k women at random in their preference list. Then the expected number of women that would benefit from lying is no more than $O(e^k)$.*

Proof: The proof will follow from a sequence of observations.

1. A woman can only benefit from a lie if she is in multiple stable matchings. Indeed if the women lies and benefits the resulting matching has to also stable. So now we will instead evaluate the probability that a woman is in more than one stable matching.

2. Women can defer their lie until they arrive at a stable matching, at which point they can discard their current partner. This will create a cascade effect (by running the stable matching algorithm again) and for her to be better off she has to get at least one other offer (since being single is the worst outcome for women). So we now only have to evaluate the probability that the woman will get a second offer after discarding her partner at the end of the run of the matching algorithm.

3. Let us analyze the run of the algorithm once a certain woman w discards his partner. What will happen is that the free man will make an offer, which will either be rejected or a new man will be freed. The process can stop in three ways.

   - w gets an offer (this is the good scenario for w)
   - The current available man has no more offers and chooses to be single.
   - A single woman gets an offer.

   It is sufficient for us to prove that the last event in the above listing is a high probability event relative to the first one, which would imply that w doesn't have much of a chance of getting another offer. So for now ignore the second event. The probability that w will be offered before any single women is $\frac{1}{|S|+1}$ where S is the set of single women (recall that the lists are randomly generated). On the other hand the probability that a woman is single is at least the probability that she is on nobody's list, and since the lists are generated randomly the latter is at most

$$(1 - \frac{1}{n})^{nk} \approx e^{-k}. \tag{1}$$

And therefore the expected number of single women is at least ne^{-k}. So therefore the probability that w benefits from lying is no more than the probability that she gets a second offer, which in turn is at most $\frac{1}{ne^{-k}}$. Thus the expected number of women that could benefit from lying is at most $\frac{n}{ne^{-k}} = e^{k}$ - a quantity that depends only on k. Thus the proof is complete.