The Stable Marriage Theorem

System of Preferences: If G is a graph, a *system of preferences* for G is a family $\{>_v\}_{v \in V(G)}$ so that each $>_v$ is a linear ordering of $N(v)$. If $u, u' \in N(v)$ and $u >_v u'$, we say that v *prefers* u to u'.

Marriage Systems and Stable Marriages: A *Marriage System* consists of a bipartite graph with bipartition $(men, women)$ which is equipped with a system of preferences. We say that a matching M is *stable* if there do not exist edges $mw, m'w' \in M$ with $m, m' \in men$ and $w, w' \in women$ so that m prefers w' to w and w' prefers m to m'.

Gale-Shapley Algorithm:

input: A marriage system on $K_{n,n}$.

output: A stable perfect matching.

procedure: At each step, every man proposes to the woman he prefers most among those who have not yet rejected him. If every woman receives at most one proposal, stop and output the corresponding matching. Otherwise, every woman who receives more than one proposal says "maybe" to the man who proposes to her whom she most prefers, and rejects the others who proposed.

Theorem 1 *The Gale-Shapley Algorithm outputs a stable perfect matching.*

Proof: Note first that this algorithm must terminate, since some man is rejected at each non-final step (and the total number of rejections is no more than n^2). Let M be the marriage resulting from this algorithm, and say that a man m and woman w are *married* if $mw \in M$.

Suppose that the woman w receives proposals from some nonempty set $X \subseteq men$ at some step. Then w says "maybe" to the man m who she prefers most among X, and at the next step, m will again propose to w (since the set of women who have rejected him has not changed). This immediately implies the following claim.

Claim: Every woman w is married to the man m she most prefers among those who propose to her during the algorithm. In particular, if w has at least one proposal, then w is married to some man.

With this claim, we now show that M covers every vertex. Suppose (for a contradiction) that M does not cover some man m. Then m must have been rejected by every woman.

But then, by the claim, every woman must be married. Since $|men| = |women|$, this is contradictory.

Next let us show that M is stable. Suppose (for a contradiction) that it is not, and choose $mw, m'w' \in M$ so that m prefers w' to w and w' prefers m to m'. It follows from the definition of the algorithm that m must have proposed to w' at some step (since m will propose to w' before w). Applying the claim to w', we see that w' must be married to m or a person she prefers to m, thus contradicting our assumptions.

It follows that M is a stable perfect matching, as claimed. $\square$

Fact: Let M be the stable perfect matching output by the above algorithm and let M' be another stable perfect matching. Then, for every man m, if $mw \in M$ and $mw' \in M'$, then either $w = w'$ or m prefers w to w'. Similarly, for every woman w, if $wm \in M$ and $wm' \in M'$, then either $m = m'$, or w prefers m' to m. So, among all stable perfect matchings, the Gale-Shapley algorithm produces one which is best possible for every man, and worst possible for every woman.