

CS261 Winter 2018 - 2019

Lecture 5: Gale-Shapley Algorithm

Instructor: Ashish Goel

Scribe: Kaidi Yan

Edited: Geoff Ramseyer

January 21, 2019

1 Problem Definition¹

We are going to study a classic problem, called the *stable marriage problem*. This is used in school choice, in resident matching, in network routers, and is one of the most beautiful matching problems. It is also something that has an anthropomorphic explanation, which we will use in explaining the problem. Please be careful in applying it to your own life — the theorems hold for very stylized cases, and may not directly translate to a complex human preference and behavior.

Consider two sets of elements V and W , where $|V| = |W| = n$. We are interested in matching each element of V with a distinct element of W . Different from the usual bipartite matching, each element of V and W has its own preference of matching. For instance, $v_1 \in V$ prefers matching with $w_1 \in W$ to matching with $w_2 \in W$, while $w_3 \in W$ prefers matching with $v_4 \in V$ to matching with $v_3 \in V$.

Our goal is to find a perfect matching between V and W such that (v, w) is *stable* for every $v \in V$ and $w \in W$. Formally, a pair (v, w) is *stable* if at least one of the following conditions is true:

1. (v, w) is paired in the matching
2. v prefers its current match to w
3. w prefers its current match to v

Equivalently, a pair (v, w) is *unstable* if v prefers w to its current match and w prefers v to its current match.

¹For more information on this problem, see The Stable Marriage Problem: Structure and Algorithms by Gusfield and Irving, and Donald Knuth's monograph at <https://www-cs-faculty.stanford.edu/~knuth/ms.html>

This problem has historically been called the “stable marriage problem” because it models the scenario where some number of men are looking to marry some number of women, and each person has some preferences as to with whom they would like to get married. Similarly, one can model the case where medical students submit rankings of where they would like to do their residency, and hospitals submit rankings of which residents they would like to take on. Many other economic market matching problems fall under this paradigm.

In fact, work on this resident-hospital matching problem and school matching problem won Lloyd Shapley (along with Stanford professor Alvin Roth) the Nobel Prize in Economics in 2012 (sadly, David Gale had passed away a few years before).

As an aside, Gale and Shapley, in their 1960 paper, give the following discourse on the mathematical elegance of this problem and their result:²

“Most mathematicians at one time or another have probably found themselves in the position of trying to refute the notion that they are people with ‘a head for figures’, or that they ‘know a lot of formulas’. At such times it may be convenient to have an illustration at hand to show that mathematics need not be concerned with figures, either numerical or geometrical. For this purpose we recommend the statement and proof of our Theorem 1 [There always exists a stable matching]. The argument is carried out not in mathematical symbols but in ordinary English, there are no obscure or technical terms. Knowledge of calculus is not presupposed. In fact one hardly needs to know how to count. Yet any mathematician will immediately recognize the argument as mathematical”

To demonstrate what Gale and Shapley meant, we will describe the algorithm in ordinary English, with minimal technical terms.

Consider the following example problem (this also illustrates why this problem has historically been called the stable matching problem): Suppose V contains three women: Alice (A), Barbara (B) and Cathy (C), while W contains three men: Dave (D), Ed (E) and Frank (F). Each man is going to marry a woman (matching between V and W), and the preference of each person’s marriage choices is given as below:

$$\begin{aligned} A : D, E, F; \quad B : D, E, F; \quad C : D, F, E \\ D : B, A, C; \quad E : B, A, C; \quad F : B, A, C \end{aligned}$$

As an explanation, list $A : D, E, F$ means that Alice prefers getting married with Dave the most, then with Ed, and finally with Frank. The rest are similar.

Here’s a possible marriage matching - $(B, D), (A, F), (C, E)$, which means that Barbara marries Dave, Alice marries Frank and Cathy marries Ed. However, this is not a stable

²Full text at <https://apps.dtic.mil/dtic/tr/fulltext/u2/251958.pdf>

matching — consider (A, E) . A and E are not paired in the matching, but A prefers E more than her current match F , and E prefers A more than his current match C . Thus (A, E) is not stable, and the matching given above is not a stable matching.

Unlike maximum bipartite matching or minimum spanning tree or many other problems you have seen before (e.g. sorting), there is no apriori reason to believe that such a matching always exists. We will give a constructive proof, i.e. prove that a stable matching always exists by providing an algorithm to find it.

2 Gale-Shapley Algorithm

The Gale-Shapley algorithm is designed to solve the stable matching problem. Again, we will describe its operation in nontechnical English

Gale-Shapley Algorithm

initialization: every man is unengaged

repeat

1. In the morning, every unengaged man proposes to the woman highest in his ranking who hasn't already rejected him
2. In the afternoon, every woman who received one or more proposals accepts the man highest in her ranking and rejects everyone else (including the one she was previously engaged to if necessary)

until: every man is engaged; then the engagements all turn into marriages and the process terminates

As a demonstration of the algorithm, consider again the example given in the first section with a different set of preference lists:

$$\begin{aligned} A : D, E, F; \quad B : D, E, F; \quad C : D, E, F \\ D : B, A, C; \quad E : B, A, C; \quad F : A, B, C \end{aligned}$$

In the first morning, D and E both propose to B ; F proposes to A .

In the first afternoon, B accepts D and rejects E ; F accepts A . Now we have (B, D) and (A, F) matched.

In the second morning, E is the only unengaged man. He proposes to A since A is the second highest in his ranking.

In the second afternoon, A accepts E and rejects F . Even if A was previously engaged to F , she stills accepts E since E is higher than F in her ranking. Now we have matchings (B, D) and (A, E) .

In the third morning, F is the only unengaged man. He proposes to B .

In the third afternoon, B accepts D and rejects F since D is higher than F in B 's ranking. Thus F remains unengaged.

In the fourth morning, F proposes to C .

In the fourth afternoon, C accepts F .

After four days, we have matchings (B, D) , (A, E) and (C, F) . We can verify that this matching is stable.

The Gale-Shapley algorithm goes on every day until every man gets engaged to a woman. Notice that this algorithm can iterate for at most n^2 days since each of the n men can propose to each of the n women at most once, and every day there's at least one man proposing to one woman (otherwise the algorithm ends). Furthermore we can restrict the bound to $n^2 - n + 1$ days as all n men propose on the first day.

This algorithm guarantees that everyone is engaged when the algorithm terminates. To see why, we argue that there cannot be a man w and a woman v both unengaged. The reason is that w must have proposed to v some day, and once v gets a proposal she chooses some man to get engaged to and after that v will remain engaged to some man (even if this man can change during different days).

Informally, we prove correctness as follows: consider a pair (v, w) when the algorithm terminates. If v and w are paired in the matching (they are engaged when the algorithm terminates) then clearly (v, w) is stable. Otherwise there are two possibilities:

1. Man w never proposes to v — Since w always proposes to the woman highest in his ranking, the fact that w never proposes to v implies that w must prefer his current match to v . Hence (v, w) is stable.
2. Woman v rejects w in favor of someone else — v rejects w if and only if v prefers some other man to w , which implies that v must prefer her current match to w . Hence (v, w) is stable.

In both cases we show that (v, w) is stable. Thus Gale-Shapley algorithm provides a stable matching in the end. Combined with the fact that Gale-Shapley must terminate with everyone getting engaged after at most $n^2 - n + 1$ days, we conclude that a stable matching always exists for any two equal-size sets with any preference lists.

In this version of the algorithm, where men do the proposing, we get a man-optimal matching, i.e., every man is at least as well off as in any other stable matching. However, it is possible that the matching it produces will not be optimal for the women. Of course, if women do the proposing, then the algorithm will give a woman-optimal matching. Game-theoretical analysis of matching problems like this one is an open area of research.