Chapter 31

Stable Matching

The Sveriges Riksbank Prize in Economic Sciences, in Memory of Alfred Nobel, for the year 2012 was awarded to Lloyd S. Shapley and Alvin E. Roth. The prize is in recognition of their pioneering contributions to theory of stable allocations and the practice of matching market design. The theory of matching and matching algorithms involves use of concepts in equal measure from non-cooperative game theory, cooperative game theory, and mechanism design. In this chapter, we describe matching algorithms and bring out their connection to game theory.

31.1 The Matching Problem

One of the common problems encountered in real life is that of matching, which is the process of allocating one set of resources or individuals to another set of resources or individuals. Common examples include matching buyers to sellers in a market; matching resources to tasks; allocating home seekers to houses; matching new doctors to hospitals; matching students to schools; matching job-seeking engineers to companies; matching advertisers to sponsored slots on a search engine page, etc. There are also examples with deep societal impact such as matching kidneys or human organs to needy patients. The matching has to be accomplished in a way that the preferences that the individuals may have, are honored, and the social welfare (measured in a reasonable way) is maximized. On the face of it, the problem looks deceptively simple, however, when the number of individuals/resources involved is large and in addition, certain inevitable, practical constraints have to be satisfied, the problem becomes complex and finding even feasible solutions (let alone optimal solutions) could be hard.

A key requirement of any solution to the matching problem is *stability*. Informally, a solution is stable if the solution cannot be improved upon, through a reallocation or further trading. Shapley and Roth have brilliantly pioneered the research and practice of the matching problem in complementary ways: In the 1960s, Shapley investigated the deep questions underlying the theory of matching and came up with an extremely elegant theory for solving the problem while Roth, in the 1980s, discovered a creative opportunity for exploiting the abstract theory of Shapley to practical problems and came up with masterly implementations to several practical problems waiting for better solutions. In the 1990s and beyond, several non-trivial extensions to the matching theory were proposed by Roth and other researchers to take into account practical issues such as strategic manipulation by the users of the matching market. Matching theory and matching markets currently constitute a lively and active area of research not only in economics but also in computer science, Internet advertising, and social computing.

Alvin Elliot Roth was born on December 19, 1951. His first degree was in operations research at the Columbia University. He completed his M.S. in 1973 and Ph.D. in 1974 both in Stanford University again in operations research. During 1975-82, he taught at the University of Illinois and he next taught at the University of Pittsburgh during 1982-98. In 1998, he joined the Harvard University where he is currently Gund Professor of Economics and Business Administration Emeritus in the Harvard Business School. Since 2013, he is Craig and Susan McCaw Professor of Economics at Stanford University.

Roth is a recipient of numerous honors including Alfred P. Sloan Fellowship, Guggenheim Fellowship, Fellowship of the American Academy of Arts and Sciences, and membership of the National Bureau of Economic Research and the Econometric Society. Besides matching markets [1], Roth has made pioneering contributions to game theory and experimental economics. One of his important contributions was to characterize the Shapley value as a risk-neutral utility function on the space of cooperative games. Roth has made influential contributions to experimental economics which are well explained in the scientific background document compiled by the Nobel Prize committee [2]. Roth's work essentially shows that the explanatory and predictive power of game theory can be enhanced with carefully and skilfully designed economic and laboratory experiments.

31.2 Matching Algorithms

In this section, we mainly discuss the college admissions problem for which Gale and Shapley [3] designed their famous algorithm.

The College Admissions Problem

Consider that there are m colleges to which a population of n students would like to get admitted (typically, $n \geq m$). Each college has a certain upper bound or quota on the number of students who could be offered admission to that college. Each college also has a strict ranking (also called preference order) of students. Likewise, each student has a strict ranking of colleges. The problem is to determine an allocation of students to colleges, according to one or more well defined criteria (such as stability, optimality, incentive compatibility, etc.). We use the words allocation, assignment, and matching synonymously.

An Example of College Admissions

Suppose there are five students, call them $1, 2, 3, 4, 5$ and two colleges A, B. Suppose the preferences of the students are as follows.

$$1 : B \succ A;\ \ 2 : A \succ B;\ \ 3 : A \succ B;\ \ 4 : B \succ A;\ \ 5 : B \succ A;$$

Let the preferences of the colleges be the following.

$$A : 1 \succ 2 \succ 3 \succ 4 \succ 5;\ \ B : 1 \succ 5 \succ 4 \succ 3 \succ 2$$

Assume that each college can take up to 3 students. An allocation would be represented by a set of pairs such as

$$\alpha_1 = \{(1, A), (2, B), (3, A), (4, B), (5, A)\}$$

Another allocation would be:

$$\alpha_2 = \{(1, B), (2, A), (3, A), (4, B), (5, B)\}$$

Yet another allocation would be:

$$\alpha_3 = \{(1, A), (2, A), (3, A), (4, B), (5, B)\}$$

Stability, Optimality, and Incentive Compatibility

We now discuss three important properties to be satisfied by a matching algorithm for college admissions: stability, optimality, and incentive compatibility.

Definition 31.1 (Stable Allocation). *An allocation is said to be* unstable *if it contains the pairs* $(1, A)$ *and* $(2, B)$ *even though* 1 *prefers* B *to* A *and* B *prefers* 1 *to* 2 *(where* $1, 2$ *are representative students and* A, B *are representative colleges). An allocation that is not unstable is said to be stable.*

Note in the above definition that if we allocate 1 to B, then both 1 and B are strictly better off. In the language of cooperative game theory, the two element coalition $\{1, B\}$ *blocks* the allocation in question. A stable allocation cannot be blocked by any coalition of size 2. The set of all stable allocations is the *core* of an underlying allocation game. The allocation game here is actually a non-transferable utility (NTU) game.

We would be interested in a stable allocation in which all students as well as all colleges are as well off under it as under any other stable allocation. That is, we seek a stable allocation such that every student (college) is matched to a college (student) that is at least as much preferred as any college (student) to which the student (college) is matched in any other stable allocation. This leads to the following definition.

Definition 31.2 (Optimal Allocation). *A stable allocation is said to be* student optimal *if every student is at least as well off under it as under any other stable allocation. A stable allocation is said to be* college optimal *if every college is at least as well off under it as under any other stable allocation.*

A student optimal allocation need not be college optimal and vice-versa. Also, it can be shown that a student (college) optimal allocation, if one exists, is unique.

Definition 31.3 (Incentive Compatible Allocation). *A stable allocation is said to be* incentive compatible for students *(colleges) if it is a best response for every student (college) to report his/her true ranking of colleges (students).*

Recall that the stronger version of incentive compatibility which is dominant strategy incentive compatibility (DSIC) in this context would mean that reporting truth is a best response regardless of what is reported by the other students (other colleges).

The Marriage Problem

The marriage problem is a special case of the college admissions problem where the number of students is equal to the number of colleges (that is $n = m$). For the sake of simplicity, we will study this problem and it turns out that many results can be generalized in a natural way to the college admissions problem. In the rest of the section, we assume that there are n students and n colleges and each student has to be matched to a college; each college can only be matched to at most one student. We shall use the numbers $1, 2, \ldots$ to denote students and the upper case letters $A, B, \ldots$ to denote colleges.

Example 31.1 (A Marriage Problem). This example is a relabeled version of the example presented by Gale and Shapley [3]. Consider three students $1, 2, 3$ and three colleges A, B, C with the following preferences.

$$1 : A \succ B \succ C;\;\; 2 : B \succ C \succ A;\;\; 3 : C \succ A \succ B$$

$$A : 2 \succ 3 \succ 1;\;\; B : 3 \succ 1 \succ 2;\;\; C : 1 \succ 2 \succ 3$$

For this situation, there are clearly six matchings or allocations. These are given by

$$\{(1, A), (2, B), (3, C)\}, \{(1, A), (2, C), (3, B)\}, \{(1, B), (2, A), (3, C)\}$$

$$\{(1, B), (2, C), (3, A)\}, \{(1, C), (2, A), (3, B)\}, \{(1, C), (2, B), (3, A)\}$$

Of these six matchings, the following can be seen to be stable:

$$\{(1, A), (2, B), (3, C)\}, \{(1, B), (2, C), (3, A)\}, \{(1, C), (2, A), (3, B)\}$$

The allocation $\{(1, A), (2, B), (3, C)\}$ can be verified to be student optimal while the allocation $\{(1, C), (2, A), (3, B)\}$ can be seen to be college optimal. The allocation

$\{(1, C), (2, B), (3, A)\}$ is neither student optimal nor college optimal. The following allocations are unstable:

$$\{(1, A), (2, C), (3, B)\}, \{(1, B), (2, A), (3, C)\}, \{(1, C), (2, B), (3, A)\}$$

To see why the allocation $\{(1, A), (2, C), (3, B)\}$ is unstable, notice that the coalition $\{3, A\}$ blocks this allocation: (a) $(3, A)$ is better than $(3, B)$ for student 3 since student 3 prefers A to B. (b) $(3, A)$ is better than $(1, A)$ for college A since college A prefers student 3 to student 1. □

Deferred Acceptance Algorithm for the Marriage Problem

We are now in a position to describe the algorithm proposed by Gale and Shapley [3] for the marriage problem. This is called the *deferred acceptance* algorithm for reasons that will become clear soon. There are two versions of this algorithm:

(1) Students-proposing version
(2) Colleges-proposing version

We shall describe the students-proposing version. The algorithm proceeds iteratively in stages.

(1) In stage 1, each student proposes to her most favored college. It is possible that some colleges may not receive any proposals; other colleges receive one or more proposals. Each college receiving two or more proposals rejects all but the most favored student. This most favored student is put in a queue (admission is deferred in the hope of getting a more favored student in a later round).
(2) In stage 2, each rejected candidate proposes to her next most favored college. Again, each college receiving two or more proposals (including any in its queue) rejects all but the most favored student. The selection of the most favored student by a college takes into account the student who might be in its queue.
(3) Stages 3, 4, etc., proceed in a manner identical to stage 2.

The algorithm terminates when every college has received a proposal. The colleges-proposing version is identical to that of the students-proposing version, except that the roles of students and colleges are reversed.

Example 31.2 (Deferred Acceptance Algorithm). This example is a relabeled version of the example presented in [2]. Consider four students and four colleges with the following preferences.

$$1 : A \succ B \succ C \succ D; \quad 2 : A \succ C \succ B \succ D$$

$$3 : A \succ B \succ D \succ C; \quad 4 : C \succ D \succ B \succ A$$

$$A : 4 \succ 3 \succ 2 \succ 1; \quad B : 4 \succ 1 \succ 3 \succ 2$$

$$C : 1 \succ 2 \succ 4 \succ 3; \quad D : 2 \succ 1 \succ 4 \succ 3$$

Suppose students propose. Then in stage 1, students $1, 2, 3$ propose to college A since this college is most favored by all of them while student 4 proposes to college C and college C puts student 4 in queue. College A rejects students $1, 2$ and puts student 3 in queue. In stage 2, the rejected students $1, 2$ propose to the next favored colleges B, C, respectively. Now, college C has one fresh proposal from student 2 and also has student 4 in queue. C now rejects student 4 and puts student 2 in queue. Also, college B places student 1 in its queue. Thereafter, in stage 3, student 4 proposes to college D. Now colleges A, B, C, and D all have one proposal each and the algorithm terminates with the allocation $\{(1, B), (2, C), (3, A), (4, D)\}$. This can be verified to be a student optimal allocation.

In the colleges-proposing version, it is easy to see that the algorithm returns the allocation $\{(1, C), (2, D), (3, A), (4, B)\}$. This can be verified to be a college optimal allocation.

Suppose the students and colleges are strategic and may not report their preferences truthfully. As an illustrative example, suppose student 4 is strategic while all other students and all the colleges are truthful. Assume that student 4 misreports her preferences as $C \succ D \succ A \succ B$ instead of her actual preferences $C \succ D \succ B \succ A$. If we now apply the colleges-proposing deferred acceptance algorithm with only the preferences of student 4 changed, then we end up with the allocation $\{(1, B), (2, C), (3, A), (4, D)\}$ which is very different from $\{(1, C), (2, D), (3, A), (4, B)\}$ that was obtained with a truthful report from student 4. In fact, this one report has turned a college optimal allocation into a student optimal allocation! This shows that the colleges-proposing version of the Gale-Shapley algorithm is not incentive compatible to the students. However, it has been shown that the colleges-proposing version is dominant strategy incentive compatible for colleges. Similarly, the students-proposing version is DSIC for students but is not incentive compatible for colleges. □

Key Properties of the Deferred Acceptance Algorithm

We now state, without proof, several attractive properties of the deferred acceptance algorithm.

- In the students-proposing (colleges-proposing) version, no student (college) proposes to the same college (student) more than once.
- Each version of the algorithm results in a unique stable allocation. The students-proposing (colleges-proposing) version results in a student (college) optimal allocation. The original paper by Gale and Shapley [3] can be consulted for an elegant proof of optimality of allocation.
- The maximum number of stages before the termination of the algorithm is $n^2 - 2n + 2$. Thus the Gale-Shapley algorithm has worst case quadratic time complexity.
- The students-proposing (colleges-proposing) algorithm is DSIC for students (colleges) but is not incentive compatible for colleges (students).
- The algorithm can be generalized in a natural way to the college admissions problem.
- Shapley and Shubik [4] have shown that the core of the assignment game (the

non-transferable utility game underlying the college admissions problem) is non-empty. This result guarantees the existence of a stable allocation in a fairly general setting.

House Allocation Problem

The deferred acceptance algorithm was a path breaking effort in the area of two sided matching problems, where the players on either side have preferences over the players on the other side. There is another important class of problems called the *house allocation* problems. In this class of problems, we have a set of players who exchange indivisible objects without any side payments. The players have preferences over objects and also own certain objects. An example situation is provided by a set of home seekers who are in search of houses for rent and have individual preferences for houses. The houses on the other hand do not have any preferences over the individuals seeking houses. Shapley and Scarf [5], in a landmark paper, designed an algorithm (based on an earlier idea from Richard Gale) called the *top trading cycle algorithm* for obtaining a stable allocation for this class of problems. This algorithm and its variants have been applied in a wide variety of problems including the kidney exchange problem.

31.3 Matching Markets

The deferred acceptance algorithm and the top trading cycle algorithm are two key algorithms for matching which have been extensively studied and modified for practical applications. A rich variety of other algorithms have been devised for matching problems of various kinds and concurrently, deep analytical studies have been carried out to establish theoretical properties (for example, stability, optimality, and incentive compatibility) of these algorithms. Lloyd Shapley who laid the foundations of cooperative game theory in the 1950s and 1960s has, in collaboration with leading game theorists, used this theory to resolve several central questions that arise in matching [1].

Alvin Roth, in the 1980s, discovered that Shapley's theory can be applied to analyze existing markets, for example the National Residents Matching Program (NRMP), a clearing house for matching new doctors with hospitals in the USA. Through a series of empirical studies and laboratory experiments, Roth and his collaborators were able to establish that stability is the most critical requirement for the success and sustainability of any matching market. In the process of this research, Roth and his team of co-researchers were able to redesign and improve the performance of many existing practical matching markets. In the 1990s, new theoretical advances, using tools including game theory, were made by Roth and colleagues in order to take care of possible strategic manipulations by the rational users of the market. Practical implementations were extended to New York schools and

Boston schools for matching students to schools. The crowning glory was achieved when the implementations were successfully applied, in the 1990s and 2000s, to the matching of organ donors to organ recipients, in particular the kidney exchange program [2].

31.4 Summary and References

In this chapter, we have only looked at some essential but elementary aspects of matching algorithms for non-strategic agents. The topic involves use of non-cooperative game theory, cooperative game theory, and mechanism design. Our exposition has covered the following aspects.

- We have touched upon the college admissions problem, the marriage problem, and the house allocation problem. In the college admissions problems, there are m colleges and n students (typically $n \geq m$) and both colleges and students have preferences. The marriage problem is a special case of the college admissions problem with $n = m$. In the house allocation problem, players on only one side have preferences.
- We have discussed the celebrated Gale-Shapley algorithm for the college admissions problems and brought out some interesting properties satisfied by this algorithm.

The award of the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel for 2012 has certainly catapulted interest in matching theory and matching markets. The scientific background document hosted at the Nobel Prize website [2] is a rich source of information on matching theory and matching markets. The book by Roth and Sotomayor [1] is an authentic source of all key results until 1990. The paper by Gale and Shapley [3] which contains the deferred acceptance algorithm is a must read.

References

[1] A.E. Roth and M. Sotomayor. *Two Sided Matching: A Study in Game Theoretic Modeling and Analysis*. Econometric Society Monograph Series, Cambridge University Press, 1990.
[2] The Economic Sciences Prize Committee. *Stable matching: Theory, Evidence, and Practical Design - The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2012: Scientific Background*. Tech. rep. The Nobel Foundation, Stockholm, Sweden, 2012.
[3] D. Gale and L. S. Shapley. "College admissions and the stability of marriage". In: *The American Mathematical Monthly* **69**(1) (1962), pp. 9–15.
[4] Lloyd S. Shapley and Martin Shubik. "The Assignment Game I: The Core". In: *International Journal of Game Theory* **1**(1) (1971), pp. 111–130.
[5] Lloyd S. Shapley and Herbert Scarf. "On cores and indivisibility". In: *Journal of Mathematical Economics* **1** (1974), pp. 23–37.