Harvard CS 1360, Spring 2025
Instructor: Ariel Procaccia
Scribe: Kelsey Chen

Matching 3: Stable Matching

Lecture 15

1 Stable Matchings

Problem formation: Say we want to match students $S = \{s_1, s_2, ...s_n\}$ with courses $T = \{t_1, t_2, ...t_n\}$. We have:

- $\pi : S \cup T \to S \cup T$ is a matching such that for all $s \in S$ and $t \in T$, $\pi(s) = t \Leftrightarrow \pi(t) = s$
- Each $s \in S$ has a ranking σ_s over T, and each $t \in T$ has a ranking σ_t over S
- A *blocking pair* for π is $(s, t) \in S \times T$ such that $s \succ_{\sigma_t} \pi(t)$ and $t \succ_{\sigma_s} \pi(s)$, i.e., a pair (s, t) who prefer each other over the way they are matched in π.

Definition 1 (Stable Matching)**.** A bipartite matching π with two-sided preferences is a stable matching if and only if there is no blocking pair.

It is not at all obvious that there should even exist stable matchings. In fact, we will soon see a constructive proof of this existence.

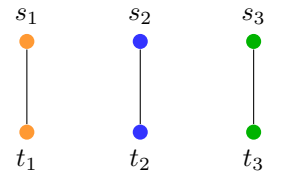
Example of a stable matching:

S Preferences

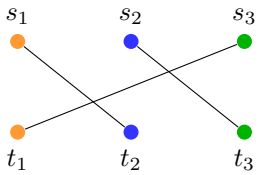
s1	s2	s3
t2	t1	t1
t1	t3	t2
t3	t2	t3

T Preferences

t1	t2	t3
s1	s3	s1
s3	s1	s3
s2	s2	s2



Unstable: (s_1, t_2) blocks



Stable

1.1 Deferred Acceptance Algorithm

The deferred-acceptance algorithm (DA) for matching with two-sided preferences takes as input a reported preference profile and outputs a matching that is stable for that profile. There are two versions of DA depending on which side makes offers. We describe the student-proposing DA, but the course-proposing DA is is defined analogously. In either version, an agent on the receiving side will tentatively hold onto the best offer they have received so far, changing their mind if something better comes along (hence "deferred" acceptance).

Definition 2 (Student-proposing deferred acceptance)**.** Each course starts off being matched with $\emptyset$. The student-proposing DA proceeds in rounds:

(Round 1) Each student, s, makes a proposal to their most preferred course. After this, each course, t, that has received a proposal tentatively accepts the proposal from their most preferred student and rejects the rest.

(Round > 1). Each student, s, whose proposal was rejected in the previous round makes a proposal to their next most preferred course. Second, each course, t, that has a new proposal tentatively accepts the proposal from their most preferred student and rejects the rest. The DA algorithm terminates when no new proposals are made, at which point the matching that corresponds to proposals that are tentatively accepted is made final.

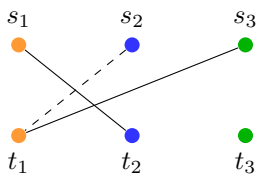
DA Example:

S Preferences

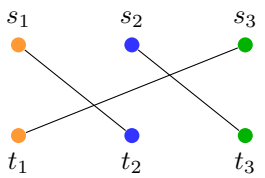
s1	s2	s3
t2	t1	t1
t1	t3	t2
t3	t2	t3

T Preferences

t1	t2	t3
s1	s3	s1
s3	s1	s3
s2	s2	s2



Round 1



Round 2

Theorem 1. *The student-proposing DA algorithm terminates with a stable matching.*

Proof. First, DA must terminate because in any round > 1 at least one proposal was rejected in the previous round and no student repeats a proposal.

To establish stability, let π denote the final matching and suppose (s, t) and (s', t') are paired, and with (s, t') blocking. Since s prefers t' to t (by def. of blocking) then s proposed to t' before t, and since t' is paired with s' then t' prefers s' to s (by course-improving across rounds). So, (s, t') does not block π. □

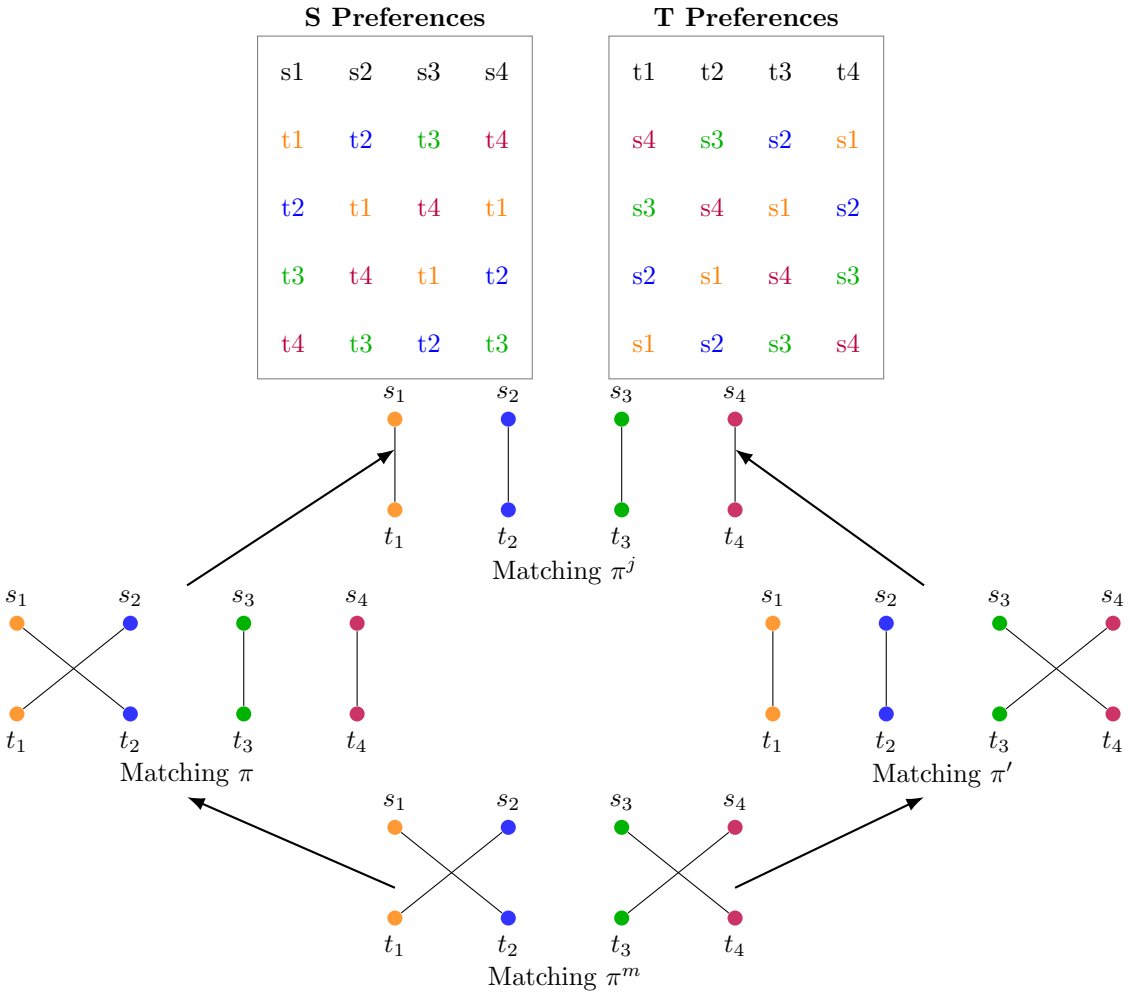
1.2 The Lattice Property

Having established existence, we now show that the set of stable matchings satisfies a remarkable property, namely that there is a stable matching that is simultaneously most-preferred by every student (and simultaneously least-preferred by every course).

- Define the *student-respecting preference ordering* $\pi \geq_S \pi'$ to mean that $\pi(s) \succeq_{\sigma_s} \pi'(s)$ for all $s \in S$ (with equality iff $\pi = \pi'$)
- Where it exists, define the *join* $\pi^j = \pi \vee \pi'$ as a stable matching π^j such that $\pi^j \geq_S \pi$, $\pi^j \geq_S \pi'$, and for every stable $\pi^\star$ satisfying these inequalities, $\pi^\star \geq_S \pi^j$. This is the least upper bound of two stable matchings π and π', from the students' perspective, i.e. it's the least student-preferred matching that is still weakly better than both π and π'
- Where it exists, define the *meet* $\pi^m = \pi \wedge \pi'$ as a stable matching π^m such that $\pi^m \leq_S \pi$, $\pi^m \leq_S \pi'$, and for every stable $\pi^\star$ satisfying these inequalities, $\pi^\star \leq_S \pi^m$ This is the greatest lower bound of two

stable matchings π and π', from the students' perspective, i.e. it's the most student-preferred matching that is weakly worse than both π and π'.

Definition 3 (Lattice Property)**.** The stable matchings are a lattice with respect to student-respecting preference ordering $\geq_S$ when the join and meet exist for any pair of stable matchings. We can think of joins as moving up the lattice and meets as moving down the lattice. Considering the join, the lattice property means that for any two stable matchings there is a stable matching that is just as preferred by every student and is smaller (in the sense of $\geq_S$) than all other improving matchings.



Theorem 2. *The meet and join exist for any pair of stable matchings.*

Proof. To prove this, define a *pointing operator* λ such that, for two matchings π and π', returns as $\lambda(s)$ whichever of $\pi(s)$ and $\pi'(s)$ is more preferred by s, and as $\lambda(t)$ whichever of $\pi(t)$ and $\pi'(t)$ is less preferred by t.

It remains to show that given two stable matchings π and π', the pointing operator λ gives a stable matching $\pi \vee_S \pi'$, the join.

- We first prove that λ is a matching:
  - Suppose for contradiction that two students s, s' where $(s \neq s')$ each choose t in λ.
  - Say (s,t) and (s',t') are matched in π and $(s,t''), (s',t)$ in π', where we might have $t' = t''$. Since s and s' both are matched to t in λ, we have $t \succ_{\sigma_s} t''$ and $t \succ_{\sigma'_s} t'$. Hence we consider t's preferences - either $s \succ_{\sigma_t} s'$ or $s' \succ_{\sigma_t} s$.

  - If $s \succ_{\sigma_t} s'$, (s,t) is blocking in π' and thus π' is not stable.
  - If $s' \succ_{\sigma_t} s$, (s',t) is blocking in π and thus π is not stable.
  - Hence $\lambda(t) = s$, i.e. each teacher is matched to exactly one student.

- Now we show λ is stable, i.e. there is no blocking pair:
  - Suppose for contradiction that (s,t) blocks matching in λ
  - W.l.o.g., suppose course t is paired in λ with $\pi(t)$, i.e. $\lambda(t) = \pi(t)$. Then $s \succ_{\sigma_t} \pi(t)$ by the definition of a blocking pair (since (s,t) is blocking).
  - Since s is paired in λ with the better of $\pi(s)$ and $\pi'(s)$, we must have $t \succ_{\sigma_s} \pi(s)$ by the definition of a blocking pair and $t \succ_{\sigma_s} \pi'(s)$
  - Hence, (s,t) is blocking in π — a contradiction in the stability of π

- Stable matching λ is the join because every matching $\pi^\star$ satisfying $\pi^\star \geq_S \pi$ and $\pi^\star \geq_S \pi'$ must at least take the student-wise max.

Hence each student points to a different course, each course to a different student, and the new matching is stable. The meet property can be established in the same way, by working with a pointing operator that returns the worst course for each student and the best student for each course. □

From this equivalence between the pointing operator and the join, we also obtain the following symmetry for the lattice of stable matchings: moving up in the lattice (join) is not only better for students but also worse for courses, and moving down (meet) is worse for students and better for courses.

It follows that there exist:

- A *student-optimal (course-pessimal) stable matching* $\bar{\pi}$ such that $\bar{\pi} \geq_S \pi$ for every stable matching π
- A *student-pessimal (course-optimal) stable matching* $\underline{\pi}$ such that $\underline{\pi} \leq_S \pi$ for every stable matching π

Theorem 3. *The student-proposing DA terminates with a student-optimal stable matching and the course-proposing DA terminates with a course-optimal stable matching*

Proof: student-proposing DA. Since students propose by order of decreasing preference, if s is matched with t, s was rejected by all t' such that $t' \succ_{\sigma_s} t$, so it suffices to show that no student is ever rejected by an *achievable course* (a course that the student could be matched with in some stable matching).

- Let R^ℓ be the set of (s,t) pairs for which t has rejected s at the start of round ℓ.
- We prove this by induction on the number of rounds ℓ, with the induction hypothesis that for every $(s,t) \in R^\ell$, t is not achievable for s. That is, in each round, no student was rejected by an achievable course. We want to show this always holds.
- The base case of $\ell = 1$ is trivial (no one has been rejected yet), so $R^1 = \emptyset$.
- In round ℓ, suppose t rejects s in favor of s', so $s' \succ_{\sigma_t} s$. We want to show that no stable matching can pair (s,t).
- Since s' proposes in order of preference, every t that s' prefers to t is already in R^ℓ at the start of the round. By the induction assumption, s' prefers t to every achievable course $t' \neq t$. Hence $t \succ_{\sigma'_s} t'$
- If there was a stable matching π with $\pi(s) = t$, $\pi(s') = t'$ for an achievable t', then (s',t) would be a blocking pair in π since $t \succ_{\sigma'_s} t'$ and $s' \succ_{\sigma_t} s$. Hence we violate the stability of π — a contradiction.

□

Similarly, the course-proposing DA outputs the course-optimal stable matching, which can be distinct from the student-optimal stable matching.

1.3 Incentives

Theorem 4. *Truthful reporting is a dominant strategy for students in the student-proposing DA mechanism.*

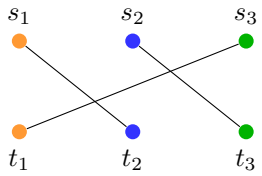
To see how courses might misreport preferences, we show an example:

S Preferences

s1	s2	s3
t2	t1	t1
t1	t3	t2
t3	t2	t3

T Preferences

t1	t2	t3
s1	s3	s1
s3	s1	s3
s2	s2	s2



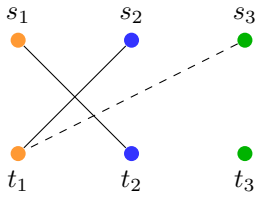
End of round 2

S Preferences

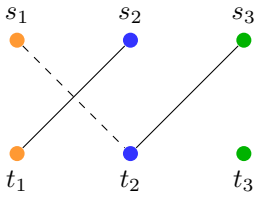
s1	s2	s3
t2	t1	t1
t1	t3	t2
t3	t2	t3

T Misreports

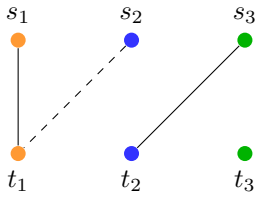
t1	t2	t3
s1	s3	s1
s2	s1	s3
s3	s2	s2



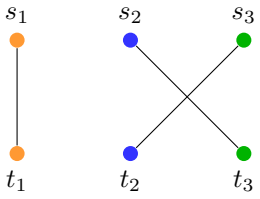
Round 1



Round 2



Round 3



Round 4

Theorem 5. *No bipartite matching mechanism with two-sided preferences is strategyproof (on both sides) and stable. (Proved in HW4)*

1.4 Stable matchings in practice

The National Resident Matching Program (NRMP) is a real-world application of the Deferred Acceptance (DA) algorithm, used to match medical school graduates (residents) to hospital residency programs in the

United States. Every year, thousands of students submit preferences over programs, and hospitals rank students. The system then computes a stable matching between residents and programs. The real NRMP uses a slightly modified version of DA:

- It is applicant-proposing (resident-proposing), which ensures the matching is optimal for residents.
- It accounts for:
  - Couples matching (where two applicants want to be matched to programs in the same city or nearby)
  - Program quotas
  - Ties or equal rankings
  - Multiple positions per program

Another important application of matching theory is in school choice, where districts use centralized systems to assign students to high schools. This is a two-sided matching problem: while students rank schools, schools also express priorities over students—often to account for factors like under-represented minorities, sibling attendance, or proximity (walk zones).

Several large school systems have implemented versions of the student-proposing Deferred Acceptance (DA) algorithm. Notably, New York City adopted such a system in 2003–04, followed by Boston Public Schools (BPS) in 2005–06. Prior to this, Boston used an immediate acceptance mechanism: students' first-choice applications were processed first, then second choices, and so on. This approach was not strategy-proof—students could benefit from misrepresenting their preferences.

When BPS switched to DA, the task force emphasized a key reason: **A strategy-proof algorithm "levels the playing field" by reducing the disadvantage faced by families who don't strategize—or who don't strategize well.** This was an explicit appeal to fairness, with strategy-proofness framed as essential for ensuring equal access, especially for families with less familiarity or comfort navigating the system. For students and families, it removes the burden of gaming the process. For policy makers, it simplifies guidance and helps gather more accurate data on school preferences to better understand demand.