# Two-sided Matching: Basics

Qianfeng Tang (SUFE)

Summer School of Modern Economics
Nanchang University, July 2015

# Marriage market

There are a set $M = \{m_1, \ldots, m_{|M|}\}$ of men and a set $W = \{w_1, \ldots, w_{|W|}\}$ of women:

- Each $m \in M$ has strict preference $\succ_m$ over $W \cup \{\emptyset\}$
- Each $w \in W$ has strict preference $\succ_w$ over $M \cup \{\emptyset\}$
- $\emptyset$ denotes "unmatched"

A **matching** is a 1-1 function $\mu : M \cup W \to M \cup W \cup \{\emptyset\}$ such that

- $\mu(m) \in W \cup \{\emptyset\}$ for all $m$
- $\mu(w) \in M \cup \{\emptyset\}$ for all $w$
- $\mu^2(i) = i$ for all $i \in M \cup W$

# Matching criterion: Stability

Suppose men and women are to be matched by a centralized mechanism, what properties should the matching satisfy? —At least they should not have incentives to divorce

## Definition

A matching $\mu$ is **stable** if it is

- Individually rational: for each $i \in M \cup W, \mu(i) \succeq_i \emptyset$
- Unblocked: there does not exist any pair $(m, w)$ such that $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$

Do stable matchings exist? If yes, how to find them?

# Example: Stable matching

Suppose $M, W$, and their preferences are

| $m_1$ | $m_2$ | $m_3$ | $w_1$ | $w_2$ | $w_3$ |
|---|---|---|---|---|---|
| $w_2$ | $w_1$ | $w_1$ | $m_1$ | $m_3$ | $m_1$ |
| $w_1$ | $w_3$ | $w_2$ | $m_3$ | $m_1$ | $m_3$ |
| $w_3$ | $w_2$ | $w_3$ | $m_2$ | $m_2$ | $m_2$ |

$\mu_1$ :

| $m_1$ | $m_2$ | $m_3$ |
|---|---|---|
| $w_1$ | $w_2$ | $w_3$ |

$\mu_2$ :

| $m_1$ | $m_2$ | $m_3$ |
|---|---|---|
| $w_1$ | $w_3$ | $w_2$ |

$\mu_1$ is stable, while $\mu_2$ is not.

# Example: The roommate problem

Four students $1, 2, 3, 4$ are to share two rooms; each room two students.

Student's preferences for roommate are:

- $\succ_1$: $2 \succ_1 3 \succ_1 4$
- $\succ_2$: $3 \succ_2 1 \succ_2 4$
- $\succ_3$: $1 \succ_3 2 \succ_3 4$
- $\succ_4$: *any*

No matching is stable: any student who is paired with 4 will and will be able to block

# Deferred acceptance algorithm (DA)

For any marriage market, the **man-proposing DA** (Gale-Shapley, 1962) operates as follows:

Step 1 Each man proposes to his most favorite woman. Each woman tentatively accepts the man that she likes the most from the proposers (if any) and rejects the rest.

Step $k, k \geq 2$ Each man rejected in the previous step proposes to his next favorite woman. Each woman tentatively accepts the man she likes the most from all proposers (tentatively accepted and new ones, if any) and rejects the rest.

Stop when no man is rejected.

# Example

The right-hand-side table below illustrates the DA procedure

| $m_1$ | $m_2$ | $m_3$ | $w_1$ | $w_2$ |
|---|---|---|---|---|
| $w_2$ | $w_1$ | $w_1$ | $m_1$ | $m_3$ |
| $w_1$ | $\emptyset$ | $w_2$ | $m_2$ | $m_1$ |
| | | | $m_3$ | |

| $w_1$ | $w_2$ | $\emptyset$ |
|---|---|---|
| $\boxed{2}, 3$ | $1$ | |
| | $1, \boxed{3}$ | |
| $2, \boxed{1}$ | | |
| | | $\boxed{2}$ |
| $\boxed{1}$ | $\boxed{3}$ | $\boxed{2}$ |

## Theorem (Gale and Shapley, 1962)

*The man-proposing DA produces the man-optimal stable matching.*

*Sketch of proof.*

- Let $\mu$ denote the DA outcome. Stability is straightforward; only need to show that $\mu$ *weakly Pareto dominates* all other stable matchings.
- Let's say that $w$ is *possible* for $m$, if $\nu(m) = w$ in some stable matching $\nu$. Given that $\mu$ itself is stable, it is sufficient to show that no man $m$ will ever be rejected by any woman $w$ who is possible for him during the DA procedure.
- Suppose not. Let step $t$ of the DA procedure be **the first step** that such rejections happen.
- Denote the man that $w$ accepts in step $t$ by $m'$, then $w$ prefers $m'$ to $m$. Consider the stable matching $\nu$ such that $\nu(m) = w$. Then $\nu(m') \neq w$. Since $\nu$ is stable, $m'$ must prefer $\nu(m')$ to $w$; otherwise $(m', w)$ will block $\nu$.
- If so, in the DA procedure, before proposing to $w$ in step $t$, $m'$ must have proposed to $\nu(m')$ (who is possible for him) and have been rejected by $\nu(m')$. This contradicts our assumption on step $t$.

# Two-sidedness: Opposite interests

Likewise, we can define **the woman-proposing DA**, which produces the woman-optimal stable matching

| $m_1$ | $m_2$ | $w_1$ | $w_2$ |
|---|---|---|---|
| $w_2$ | $w_1$ | $m_1$ | $m_2$ |
| $w_1$ | $w_2$ | $m_2$ | $m_1$ |

$\mu_1 : \dfrac{m_1 \quad m_2}{w_2 \quad w_1}$ $\mu_2 : \dfrac{m_1 \quad m_2}{w_1 \quad w_2}$

- $\mu_1$ is man-optimally (woman-worst) stable
- $\mu_2$ is woman-optimally (man-worst) stable

**Caution**: The man-optimal stable matching is not necessarily Pareto efficient for men (see previous example)

# Structure of the set of stable matchings

Let $\mu$ and $\mu'$ be two matchings. Define their **join** $\lambda = \mu \vee \mu'$ and **meet** $\nu = \mu \wedge \mu'$ by letting

- $\lambda(m) = \max_{\succ_m}\{\mu(m), \mu'(m)\}$ for all $m$, and
  $\lambda(w) = \min_{\succ_w}\{\mu(w), \mu'(w)\}$ for all $w$
- $\nu(m) = \min_{\succ_m}\{\mu(m), \mu'(m)\}$ for all $m$, and
  $\lambda(w) = \max_{\succ_w}\{\mu(w), \mu'(w)\}$ for all $w$

## Theorem (Lattice theorem, Conway)

*If both $\mu$ and $\mu'$ are stable, then both $\lambda$ and $\nu$ are matchings and both are stable.*

That is, the set of stable matchings is a lattice. A more direct approach on this result employs Tarski's fixed point theorem.

# Incentives to misreport

A (centralized) **mechanism** $\varphi$ maps each profile of agents' reported preferences $\succ \equiv ((\succ_m)_{m \in M}, (\succ_w)_{w \in W})$ to a matching $\varphi(\succ)$

## Definitions

$\varphi$ is **strategy-proof** if it is a weakly dominant strategy for every agent to report his/her preference truthfully, i.e.,

$$\varphi(\succ_i, \succ_{-i})(i) \succeq_i \varphi(\succ'_i, \succ_{-i})(i), \forall i \in M \cup W, \succ_{-i}, \succ'_i .$$

## Theorem (Roth)

*There is no stable mechanism that is strategy-proof.*

*Proof*. Revisit the example on the slide Two-sidedness. If $\varphi$ is a stable mechanism, then it has to select either $\mu_1$ or $\mu_2$ for the given preference profile. If it selects $\mu_1$ (the argument is similar if it selects $\mu_2$), then $w_1$ will have incentive to misreport $\succ'_{w_1}$: $m_1 \succ \emptyset$. At $(\succ'_{w_1}, \succ_{-w_1})$, the only stable matching is $\mu_2$, hence $\varphi$ selects $\mu_2$ and $w_1$ becomes better off.■

# The "lone wolf" theorem

The impossibility result is mainly due to the two-sidedness: when one side is happy, the other side is not and can gain from misreporting.

## Theorem (McVitie-Wilson, 1970)

*The set of matched men (women) is the same across all stable matchings.*

*Proof*. Let $\bar{\mu}$ be the man-optimal stable matching and $\mu$ be any stable matching. Then men prefer $\bar{\mu}$ to $\mu$ while women prefer $\mu$ to $\bar{\mu}$. Therefore, weakly more men is matched at $\bar{\mu}$ and weakly more women is matched at $\mu$. We also know that at each matching, the numbers of matched man and women are the same. Hence $|\bar{\mu}(M)| = |\mu(M)|$ and consequently, $\bar{\mu}(M) = \mu(M)$, where $\mu(M)$ is the set of matched men at $\mu$. ■

$$\begin{array}{ccc} \bar{\mu}(M) & \overset{\text{card}}{=} & \bar{\mu}(W) \\ \cup & & \cap \\ \mu(M) & \overset{\text{card}}{=} & \mu(W) \end{array}$$

As a direct consequence of the "lone wolf" theorem, under a stable mechanism, a man/woman can misreport to obtain any stable assignment

## Corollary

*If $\varphi$ is a stable mechanism, and $\mu$ is a stable matching at $(\succ_i, \succ_{-i})$, then there exists $\succ'_i$such that $\varphi(\succ'_i, \succ_{-i})(i) = \mu(i)$.*

*Proof*. Let $\succ'_i$: $\mu(i)\emptyset$. That is, agent $i$ misreport that only $\mu(i)$ is acceptable for him/her. Note that $\mu$ is also stable under $(\succ'_i, \succ_{-i})$. Due to the "lone wolf" theorem, $|\mu(i)| = |\varphi(\succ'_i, \succ_{-i})(i)|$. That is, if $i$ is matched at $\mu$, he/she must also be matched at $\varphi(\succ'_i, \succ_{-i})$. Since only $\mu(i)$ is acceptable to $i$ at $\succ'_i$ and $\varphi$ is stable, $\varphi(\succ'_i, \succ_{-i})(i) = \mu(i)$. ■

## Theorem (Dubins-Freeman; Roth)

*The man-proposing DA mechanism is strategy-proof for all men*

Likewise, the woman-proposing DA is strategy-proof for all women

# Application: One-sided matching

Next, we consider applying the framework of two-sided matching to school choice problems

Although the matching is still two-sided, we often view school choice problems as one-sided allocation problems, where objects (seats at schools) are to be assigned to agents (students)

The most important criteria for school-choice mechanisms are: Pareto efficiency, stability, and strategy-proofness

As a result of the one-sidedness, for efficiency, we care only about the welfare of students, and for strategy-proofness, we only care about the strategic behavior of students. Lastly, stability is interpreted as a notion of fairness among students

# School choice

In school choice, we assign seats at schools $S = \{s_1, \ldots, s_n\}$ to students $I = \{i_1, \ldots, i_m\}$:

- $q_s$ : the **capacity** (quota) of school $s$
- $\succ_s$: the strict **priority structure** at school $s$; $i \succ_s j$ iff $i$ has higher priority than $j$ at $s$
- $P_i$ : the strict **preference** of student $i$; $s' P_i s$ iff $i$ prefers $s'$ to $s$
- $R_i$ : the extension of $P_i$. That is, $s' R_i s$ iff $s' P_i s$ or $s' = s$
- $\emptyset$ : the **null school**, unlimited capacity
- $\succ \equiv (\succ_s)_{s \in S}, P \equiv (P_i)_{i \in I}$. And $\succ_{-s}, P_{-i}$ are defined as usual

A **school choice problem** is a pair $(P, \succ)$

# Matching criteria

A **matching** is a function $\mu : I \to S \cup \{\emptyset\}$ such that $|\mu^{-1}(s)| \leq q_s, \forall s$.

For a given problem $(P, \succ)$ and matching $\mu$ :

- $\mu$ is **Pareto efficient** (for students) if there is no matching $\nu \neq \mu$ such that $\nu(i) R_i \mu(i), \forall i$
- $\mu$ is **stable** if it is not blocked by any student-school pair $(i, s)$. Formally, if it is
  - **fair**: there is no $i, j, s$ such that $i$ violates $j$'s priority at school $s$, i.e., $\mu(j) = s, i$ desires $s$ $(s P_i \mu(i))$, but $i \succ_s j$; and
  - **non-wasteful**: $i$ desires $s \Rightarrow |\mu^{-1}(s)| = q_s$

In school choice, schools' priority structure $\succ$ is often exogenously given. A centralized allocation mechanism asks each student to report his/her preference and based on that, make assignments.

# Deferred acceptance algorithm

Let $P$ denote the set of all possible preference profiles and $M$ the set of matchings. An **allocation mechanism** is a mapping $\varphi : P \rightarrow M$

In particular, each given priority structure $\succ$ defines a student-proposing DA mechanism, denoted by $DA^{\succ}(\cdot)$.

For each preference profile $P$, the DA outcome $DA^{\succ}(P)$ is produced as follows:

Step 1: Each student applies to her most favorite school. Each school **tentatively** accepts the best students up to its capacity and rejects the rest.

Step $k, k \geq 2$: Each student rejected in the previous step applies to her next best school. Each school **tentatively** accepts the best students up to its capacity, by comparing both accepted students and new applicants.
Stop when no student is rejected.

# Boston mechanism

For each preference profile $P$, the outcome of the Boston mechanism is produced as follows:

Step 1 Each student applies to her most favorite school. Each school (**permanently**) accepts the best students up to its capacity and rejects the rest.

Step $k, k \geq 2$ Each student rejected in the previous step applies to her next best school. Each school that still has available seats (**permanently**) accepts the best students among the new applicants up to its remaining capacity.
Stop when no student is rejected.

Under the Boston mechanism, the acceptances are first-come-first-serve, and are permanent instead of tentative. Therefore, students have strong incentives to put their targeting school at the top of their reporting preference lists

# Example: DA and Boston

In the problem $(P, \succ)$, $S = \{s_1, s_2\}$, $q_{s_1} = q_{s_2} = 1$; $I = \{i, j, k\}$. Tables below illustrate the priority/preference and the DA and Boston Procedures

| $\succ_{s_1}$ | $\succ_{s_2}$ | $P_i$ | $P_j$ | $P_k$ |
|---|---|---|---|---|
| $i$ | $k$ | $s_2$ | $s_1$ | $s_1$ |
| $j$ | $i$ | $s_1$ | $\emptyset$ | $s_2$ |
| $k$ | | | | |

| $s_1$ | $s_2$ | $\emptyset$ |
|---|---|---|
| $\boxed{j}, k$ | $\boxed{i}$ | |
| | $\boxed{k}, i$ | |
| $\boxed{i}, j$ | | |
| | | $\boxed{j}$ |
| $\boxed{i}$ | $\boxed{k}$ | $\boxed{j}$ |

| $s_1$ | $s_2$ | $\emptyset$ |
|---|---|---|
| $\boxed{j}, k$ | $\boxed{i}$ | |
| | $\boxed{i}, k$ | |
| | | $\boxed{k}$ |
| $\boxed{j}$ | $\boxed{i}$ | $\boxed{k}$ |

- Although DA is always optimally stable and strategy-proof, for this given $\succ$, $DA^{\succ}$ is not Pareto efficient
- Boston mechanism is not stable, and is not strategy-proof (although its Pareto efficient under the truthful preferences)

# Practice: Chinese college entrance exam

Starting with 2001, the Parallel mechanism–a variation of DA–gradually replaces the Boston mechanism in Chinese college admission

Chen and Kesten (2015), "Chinese College Admissions and School Choice Reforms: Theory and Evidence" for detailed documentation of the history of the admissions reform