Networks II: Market Design—Lecture 18 Matching Markets with Non-transferable Utilities: Two-sided preferences

ARPITA GHOSH

Dept. Of Information Science,
Cornell University

Outline

- Recap of last lecture:
  - Comparing matchings: For any two **stable** matchings μ and μ', $\mu \succeq_M \mu'$ if and only $\mu' \succeq_W \mu$
    - Corollary: μ_M^* is woman-pessimal matching
  - Same set of agents is matched in all stable matchings (!)
  - Incentives
- Coming up today:
  - Incentives: One-sided strategyproofness
  - An application: The NRMP match

Incentives and stability: Recap

- An example showing agents may have incentives to manipulate DA (deferred acceptance) mechanism
- Call a mechanism a *stable matching mechanism* if it returns a stable matching with respect to *reported* preferences

Theorem (Impossibility Theorem; Roth 1982)

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

Incentives and stability: Recap

- Matchings and preferences:
  - $\mu_1 = (m_1 - w_1, m_2 - w_2)$; $\mu_2 = (m_1 - w_2, m_2 - w_1)$
  - Preferences $\succ_{m_1}$: w_1, w_2; $\succ_{m_2}$: w_2, w_1;
    $\succ_{w_1}$: m_2, m_1; $\succ_{w_2}$: m_1, m_2
- Recall: μ_1 and μ_2 are outcomes of man and woman-proposing DA respectively
  - If M returns μ_1, w_1 benefits from lying: But neither man can benefit from lying
  - If M returns μ_2, m_1 can profitably lie: Neither woman can benefit from lying
- Is there a pattern here?

One-sided incentive properties

Theorem (Dubins and Freedman 1981, Roth 1982)

The man-proposing DA algorithm is strategy-proof for men, i.e., reporting preferences truthfully is a dominant strategy *for every man (and woman-proposing DA is strategy-proof for women).*

- 'One-sided' strategy-proofness: No man can benefit by lying, *irrespective* of what all other agents (including women!) report

Why is man-proposing DA strategy-proof for men?

- An outline of the main idea:
  - Fix reports of all women and all but one man m
  - Will show: Whatever report m starts with, modifying to truthful preference reporting weakly improves partner returned by man-proposing DA
  - If so, no man m can strictly benefit from lying: Strategyproof for men!
- Let $\succ_m$ denote true preferences of m
- Suppose m reports preferences $\succ'_m$: Man-proposing DA returns matching μ', with $w' = \mu'(m)$
- Demonstrate a sequence of changes, starting with $\succ'_m$, and ending at $\succ$: Each change weakly improves outcomes

Why is man-proposing DA strategy-proof for men?

Each of the following changes weakly improves m's outcome:

1. Modify to report preference $\succ''_m = (w')$: w' is *only* acceptable woman
   - μ' still remains a stable matching (Why?)
   - μ' matches m (with w'): So m must be matched in every stable matching with these preferences (Why?)
   - So m must be matched to w' in every stable matching when m switches to report preference $\succ''_m$, including in outcome of (man-proposing) DA

Why is man-proposing DA strategy-proof for men?

2. Modify preferences to $\succ_m'''$: True preferences *truncated* at w'
   - No matching where m is unmatched can be stable
     - Suppose not: Let $\hat{\mu}$ be a stable matching with preference $\succ_m'''$ where m is unmatched
     - $\hat{\mu}$ must also be stable when m reports $\succ_m''$, with only w' in preference list (Why?)
     - Contradiction to m being matched in every stable matching under preference $\succ_m''$ argued previously
   - So m must be matched in every stable matching under report $\succ_m'''$
   - m must get a partner he likes at least as much as w' under man-proposing DA

3. A final change from $\succ'''_m$ to $\succ_m$ that also 'weakly improves' (here, does not change) m's outcome:
    - $\succ$: Reporting honestly with no truncation
    - Does not affect DA outcome relative to report $\succ'''_m$ (Why?)

- So reporting preferences $\succ_m$ can only (weakly) improve outcome relative to reporting $\succ'_m$

One-sided incentive properties

Theorem (Dubins and Freedman 1981, Roth 1982)

The man-proposing DA algorithm is strategy-proof for men, i.e., reporting preferences truthfully is a dominant strategy *for every man (and woman-proposing DA is strategy-proof for women).*

- Advanced material (we won't address this):
  - Stronger result: No *coalition* of men can (all strictly) benefit by lying in man-proposing DA
  - 'What actually happens' in a one-sided strategyproof mechanisms?
    - Analysis of *equilibrium* behavior; incentives in *large* markets

Understanding the result

- Which previous theorems did we need to prove strategyproofness for men?
  - Same set of agents matched in all stable matchings (whose proof used existence of man-optimal, woman-pessimal stable matching!)
- Which step in the proof fails for women (with the same man-proposing DA)?

Summary: Strategyproofness

- DA algorithm is not strategyproof
- **No** stable matching mechanism is strategyproof
- Man-proposing DA is strategyproof for men
- Women-proposing DA is strategyproof for women

Stable allocations with two-sided preferences

Recall from our first foray into two-sided markets:

- Immediate questions:
  - How does stability relate to Pareto-efficiency, core in two-sided markets?
  - Does a stable matching always *exist*? Can we find one if it does exist?
  - What are the properties of stable matchings?
  - Stability is theoretically appealing, but does it matter in real life?
- Coming up next: An application
  - Lightning intro to many-to-one matchings
  - Stability and the NRMP match

Many-to one matching

In many real applications, one side of the market has *multi-unit* demand

- College admissions: Colleges accept multiple students
- Medical residency/internships: Hospitals may admit more than one resident
- Labor markets: Firms want to employ multiple workers
- . . .

Many-to one matching: The College Admissions Model

- Finite sets S of students and C of colleges: **Each college c has q_c positions to fill**
- Students have rank-order preferences over colleges
- Colleges have rank-order preferences over **individual** students
- Preferences over sets of students 'derived' from preferences over individual students
  - Note: A college may not be able to compare arbitrary subsets of students in this model
  - Example: Suppose $q_c = 2$ and $\succ_c$: $s_1 \succ s_2 \succ s_3 \succ s_4$
  - Then (s_1, s_3) is preferred to (s_1, s_4), but (s_2, s_3) and (s_1, s_4) are not comparable

Many-to one markets: The College Admissions Model

- Many-to-one matching μ: Pairings of colleges with (disjoint) *sets* of students
- Formally, μ is function from $S \cup C$ to $S \cup C \cup \{\emptyset\}$ such that
  - $\mu(s) \in C \cup \{\emptyset\}$
  - $\mu(c) \subseteq S$ (Each college is matched to a *group of students*)
  - $|\mu(c)| \leq q_c$ (No college's quota is exceeded)
  - $\mu(s) = c \iff s \in \mu(c)$, for every student $s \in S$ and college $c \in C$

Many-to one matchings: Stability

- Stability: No college-student pair that prefer each other to current allocation
- Matching μ is stable in M if no (mutually acceptable) pair (s, c) such that
  - s is unmatched and c has not filled its quota q_c
  - s is unmatched and c prefers s to at least one of its students
  - c has not filled its quota q_c and s prefers c to its assigned college $\mu(s)$
  - s prefers c to $\mu(s)$, and c prefers s to *at least one* student in $\mu(c)$

Properties of many-to-one matchings

- Many results from one-to-one matching carry over to many-to-one matching:
  - Existence of stable matchings
    Proof A neater proof
  - Unmatched agents: “Rural hospital theorem”
    Proof
  - Student (respectively college)-proposing DA returns the student (respectively college)-optimal stable matchings
- But **not** all properties hold: In particular, incentive properties
  Incentives

Mechanisms in real markets: The NRMP match

- Hospital-intern medical labor market: The NIMP/NRMP match
  - NIMP/NRMP: National Intern/Residency Matching Program
  - Entry-level labor market for physicians
  - Graduating medical students interview at hospitals
  - Students, hospitals form preferences over each other
- Market evolution:
  - Decentralized market beginning 1900s
  - Unraveling, congestion around 1945: Centralization via NIMP/NRMP
  - High participation, until rates began to decline in 1970s
- Similar matching programs in other countries: Some successful, some less so

The NRMP match: Questions

- Why was (voluntary) participation so high in the NRMP match?
  - Other countries: How to understand which programs succeed?

Participation

- Rural hospitals get too few interns: Can this be fixed?

Rural hospitals

- Medical associations claim algorithm incentivizes reporting true preferences, but students disagreed: Who is right?

Incentives

- Why decline in participation rates in 1970s, and how to fix it?

Match variations

The NRMP match

- NRMP match starting 1951: Centralized market clearing
  - Stable marriage model, Deferred Acceptance algorithm: 1962 (Gale-Shapley)
  - Algorithm used for centralized assignment in NRMP: $\approx$ 1951
    NRMP Algorithm
- What are the properties of the matching chosen by the NRMP algorithm?

Theorem (Roth'84)

The NRMP algorithm returns a stable matching. Further, the NRMP algorithm produces the same outcome as the hospital-proposing deferred acceptance (DA) algorithm.

Mechanisms in real markets

How does 'success' of market relate to stability of matching?

- Roth (1984, 1991): A theoretical and empirical study of medical residency markets
    - American hospital-intern matching (NIMP) algorithm produces stable matching
    - British medical match: Different regions use different matching mechanisms
        - Stable mechanisms continue to be used
        - Most unstable mechanisms abandoned after short period of time
- Empirical evidence from other matching markets also suggests: Stability matters!

Mechanisms in real markets

Market	Stable	Still in use
American medical markets		
NRMP	yes	yes (new design 98-)
Medical Specialties	yes	yes
British medical markets		
Edinburgh ('69)	yes	yes
Cardiff	yes	yes
Birmingham	no	no
Edinburgh ('67)	no	no
Newcastle	no	no
Sheffield	no	no
Cambridge	no	yes
London Hospital	no	yes
Other healthcare markets		
Dental Residencies	yes	yes (2/7 no)
Osteopaths (-'94)	no	no
Osteopaths ('94-)	yes	yes
Other matching markets and processes		
Canadian Lawyers	yes	yes
Reform rabbis	yes	yes
NYC highschool	yes	yes

The "Rural Hospital Theorem"

- Allocation of residents in rural hospitals:
  - Hospitals in rural areas cannot fill positions for residents
  - Change matching mechanisms so that more doctors end up in rural hospitals?

Theorem

The set of students and colleges that are unmatched is the same for all stable matchings: A student unmatched in some stable matching is unmatched in every stable matching; all colleges fill the same number of positions across stable matchings.

- Theorem: Impossible as long as **stable** matching is implemented
  - Empirical studies: Stability drives participation

The "Rural Hospital Theorem"

- Also addresses fairness amongst students:
  - If some students are matched in some stable matching and not in others, the latter may be unfair to him/her: Theorem says this won't happen
- What about improving quality of doctors assigned to rural hospitals that cannot fill their quota?

Theorem

The set of students assigned to a college that does not fill its quota in a stable matching is identical across all stable matchings. (!)

Understanding NRMP

Two ingredients: Recall

- **Theorem**: The outcome of the NRMP algorithm is the same as from the hospital-proposing deferred acceptance algorithm.
- Note hospitals $\equiv$ colleges
- **Theorem**: The college-proposing DA algorithm is not strategyproof for students. Also, no stable mechanism is strategyproof **for colleges**: in particular, the college-proposing DA algorithm is **not** strategyproof for colleges.

So NRMP algorithm is not strategyproof: Students and colleges can indeed benefit by misreporting preferences

Understanding NRMP

Match variations

- Medical residency match has additional complexities not in simple model
- 'Match variations': Complex markets
  - 'Supplemental rank-order lists': Applicants seeking second-year residencies with first-year prerequisities
  - Residency programs with interdependent quotas
  - Programs that want to fill an even number of matches
  - Couples who want close-by residencies: Joint preferences
    - Stable matchings need not even exist if there are couples!

What next?

What next, when theory is inadequate for real markets?

- Design: How to modify algorithms to account for match variations?
- Empirical analysis: What is the magnitude of problems in real market?
- Theory: Why does the market nonetheless 'seem to work'?
  - Theory of large markets

Large matching markets

- Many hospitals and students, but each student can realistically interview only at a small number of hospitals
- Each student only rank-orders hospitals he has interviewed at (all other hospitals treated as unacceptable)
- Result (informal): Changes in partners, incentives for strategic manipulation vanish as markets grow large
  - Only agents who obtain different partners in college and student-proposing matchings can successfully manipulate
  - This number becomes very small as market grows large

Theorems

- NRMP case study: Matching market with two-sided preferences
- Many-to-one matching theory: Answers many questions about practice
- Theory predicts problems: What then?
  - Modify design to partially help (Roth-Peranson algorithm)
  - Empirical evaluation: Magnitude of problems is small in real markets
  - Why do markets nonetheless 'seem to work'? New theory of large markets

Supplemental material

Note: *The following material is supplemental, and not part of the course syllabus*

- Many-to-one matchings: Results, and (some) proofs
  - Existence of stable matchings
  - Unmatched agents: "Rural hospital theorem"
  - Student (respectively college)-proposing DA returns the student (respectively college)-optimal stable matchings
  - Incentive properties

Stable matchings always exist in many-to-one matching

Theorem (Gale and Shapley 1962)

There exists a stable matching in any many-to-one matching market.

- Generalize the (student-proposing) DA:

Step 1 (a) Each student "applies" to her first choice college
(b) Each college tentatively holds the most preferred (acceptable) applicants **up to its quota**, and rejects all other applicants

Step $t \geq 2$ (a) Each student rejected in Step $(t-1)$ applies to her next highest choice
(b) Each college considers both new applicants and the students (if any) held at Step (t-1), tentatively holds the most preferred acceptable students **up to its quota** from the combined set of students, and rejects all other students

- Terminate when no more applications are made

Does this produce a stable many-to-one matching?

- DA returns stable matching: Essentially identical argument
- Recall key observations for stability in one-to-one matching: For every woman w,
  1. w is always engaged after she first accepts a proposal
  2. w has an *improving* sequence of partners

  For every man m,

  3. Sequence of women to whom m proposes gets worse and worse
- Observations generalize appropriately: Stability of many-to-one matching

A *reduction* from many-to-one markets to one-to-one markets:

- M: Original many-to-one matching market
- Construct *one-to-one* market $\hat{M}$ from M:
  - Each college c is replaced by q_c different colleges $\hat{c}_i$, $i = 1, \ldots, q_c$ with one position each
  - New colleges inherit (strict) preferences of old colleges over students
  - Each student chooses an arbitrary tie-breaking order to strictly rank the copies of each college in $\hat{M}$
- Every (valid) many-to-one matching μ in M can be mapped to a matching $\hat{\mu}$ in $\hat{M}$ and vice versa

Reduction

Lemma

A matching μ is stable in the many-to-one market M if and only if the 'corresponding' matching $\hat{\mu}$ is stable in $\hat{M}$.

- If direction: Suppose $\hat{\mu}$ is stable in $\hat{M}$
  - No blocking pair can exist in μ: Suppose for contradiction that pair (s, c) blocks μ in M
  - $(s, \hat{c}_i)$ must also block $\hat{\mu}$ for some copy $\hat{c}_i$ of c (Why?)
- Only if direction: Suppose μ is stable in M
  - No blocking pair can exist in $\hat{\mu}$: If $(s, \hat{c}_i)$ blocks $\hat{\mu}$, (s, c) must block μ as well
  - μ could not be stable, a contradiction

Many-to-one stable matchings

Theorem (Gale and Shapley 1962)

There exists a stable matching in any many-to-one matching market.

A neater argument:

- Replace every college with quota q_c with q_c new colleges each with quota 1
- Run deferred acceptance algorithm in this new market
- What does this argument involve?
  - Define preferences of students over new 'colleges' (Careful: Strict preferences!)
  - Stability in original market iff stability in new market

Properties of stable many-to-one matchings

Theorem ("Rural hospital theorem")

A student is either unmatched in all stable matchings or matched in all stable matchings; the number of students assigned to a given college is the same across all stable matchings.

- Reduction: Stable μ in M corresponds to stable $\hat{\mu}$ in $\hat{M}$ (and vice versa)
- Set of unmatched agents same in all stable matchings in $\hat{M}$:
  - If s is matched in stable matching $\hat{\mu}$, s also matched in corresponding (stable) matching μ in M
  - Number of positions filled is c same in all stable matchings μ in M
    - Each $\hat{c}_i$ is either matched in all stable matchings $\hat{\mu}$ in $\hat{M}$ or unmatched in all stable matchings
    - Number of filled positions for c in μ: Number of copies $\hat{c}_i$ matched in $\hat{\mu}$

Incentives in many-to-one matchings

Theorem

No *stable mechanism is strategyproof* **for colleges***: In particular, the college-proposing DA algorithm is* **not** *strategyproof for colleges.*
However, the student-proposing DA algorithm remains strategyproof for students.

- Note contrast with one-to-one matching in marriage model: *If* every college had quota exactly 1, college-proposing DA would be strategyproof for colleges
- A rule of thumb: Reduction-based arguments often don't carry over for incentives!

Couples and large markets

Formal statements from Kojima, Pathak, Roth (QJE 2013): Large market, small number of couples.

Theorem

The probability that there exists a stable matching converges to one, as the size of the market (number of colleges) goes to infinity with the number of couples being fixed.

Theorem

For any $\varepsilon > 0$, there exists n such that truth-telling by every agent is an ε-Nash equilibrium under the Roth-Peranson algorithm for any game with more than n colleges.