Introduction to Market Design and Two-Sided Matching Theory

Fuhito Kojima

April 3, 2018

Some Logistical Issues

1. Textbook: *Two-Sided Matching* by Roth and Sotomayor (1990). I will also cover recent journal articles and working papers.
   1. A more recent survey papers. Roth (2008), Sonmez and Unver (2009), Abulkadiroglu and Sonmez (2013), Kojima (2017), Pathak (2017).
   2. A general audience book by Roth (2015). Fun to read.
2. Office Hours: by appointment (email me)
3. 3 problem sets (tentative plan: posted on April 5th, 19th, and May 3rd), one or more class presentations *and participation*, one final paper (5-8 pages).
4. No coding required, although algorithms are central (see David Manlove's book)
5. Possible class plan change

Some Logistical Issues

1. We use math and formal arguments, though motivated by real-life issues.
2. Not much advance math knowledge is needed, but familiarity with (and willingness to engage in) formal mathematical arguments is essential.
3. A good complement:
   - Al Roth's market design blog: http://marketdesigner.blogspot.com/
   - 2012 Nobel Prize, http://www.nobelprize.org/nobel_prizes/economics/laureates/2012/
4. Stanford is unusually strong in market design: Al Roth, Paul Milgrom, Muriel Niederle, Jon Levin, Ilya Segal, Gabriel Carroll (Econ department), Michael Ostrovsky, Susan Athey, Mohammad Akbarpour, Gabriel Weintraub, Daniela Saban (GSB), Ramesh Johari, Tim Roughgarden (CS), Itai Ashlagi (MS&E) and many others ...

What is Matching and Market Design?

- Recently economists have been using economics to design institutions successfully, such as (1) student placement in schools, (2) labor markets where workers and firms are matched, and (3) organizing organ donation network, and many more are being proposed and implemented, such as (4) course allocation in schools, (5) cadets to branches in the military, and so on.
- The economics of "matching and market design" analyzes and designs real-life institutions. A lot of emphasis is placed on concrete markets and details so that we can offer practical solutions.

Labor Markets: The case of American hospital-intern markets.

- Medical students in many countries work as residents (interns) at hospitals.
- In the U.S. more than 20,000 medical students and 4,000 hospitals are matched through a clearinghouse, called NRMP (National Resident Matching Program).
- Doctors and hospitals submit preference rankings to the clearinghouse, and the clearinghouse uses a specified rule (computer program) to decide who works where.
- Some markets succeeded while others failed. What is a "good way" to match doctors and hospitals?

Kidney Exchange

- Kidney transplant is a preferred method to save kidney-disease patients.
- There are lots of kidney shortages, and willing donor may be incompatible with the donor.
- Kidney Exchange tries to solve this by matching donor-patient pairs.
- What is a “good way” to match donor-patient pairs?

School Choice

- In many countries, especially in the past, children were automatically sent to a school in their neighborhoods.
- Recently, more and more cities in the United States and in other countries employ **school choice** programs: school authorities take into account preferences of children and their parents.

School Choice

- Because school seats are limited (for popular schools), school districts should decide who is admitted.
- How should school districts decide placements of students in schools?
  - Typical goals of school authorities are: (1) efficient placement, (2) fairness of outcomes, (3) easy for participants to understand and use, etc.
- To study these questions (and others), we will study the theory of matching beginning today.

A simple theory of matching

- Proposed by Gale and Shapley (1962). Suggested reading is Roth and Sotomayor (RS henceforth), Chapter 2.
- Finite sets S of students and C of colleges (we use student-college terminology just for convenience).
- Each student can be matched to at most one college, and each college can admit at most one student (so the model is called "one-to-one matching").
    - Students have strict preferences over colleges and being unmatched (denoted by $\emptyset$) and colleges have strict preferences over students and being unmatched.
- $c \succ_s c'$ means "student s strictly prefers college c to college c'. $s \succ_c s'$ means "college c strictly prefers student s to student s'.
- If $i \succ_j \emptyset$ then we say i is **acceptable** to j.

Matching

- The outcome of the matching market is a **matching**, which specifies which student attends which college (if any).
- Formally, matching μ is a function from $S \cup C$ to $S \cup C \cup \{\emptyset\}$ such that
    1. $\mu(s) \in C \cup \{\emptyset\}$ (i.e. a student is either matched to a college or unmatched),
    2. $\mu(c) \in S \cup \{\emptyset\}$ (i.e. a college is either matched to a student or unmatched), and
    3. $\mu(s) = c \iff \mu(c) = s$, for every student $s \in S$ and college $c \in C$ (i.e., matching is mutual -"you are matched with me if and only if I am matched with you").

Stability

- The most important desideratum of a matching is called stability.
- Roughly speaking, a matching is stable if there are no individual players or pairs of players who can profitably deviate from (block) it.

Stability

- Formally, stability is defined as follows.
- Matching μ is **blocked by an individual** i (aka **individually irrational**) if $\mu(i)$ is unacceptable to i, that is, $\emptyset \succ_i \mu(i)$.
- Matching μ is **blocked by a pair** s and c if each of them prefers each other to their partners under μ, that is,

$$c \succ_s \mu(s), \text{ and } s \succ_c \mu(c).$$

- A matching is **stable** if it is not blocked by any individual or pair.
- (a note: the set of all stable matchings is equivalent to the **core**, and a stable matching is **Pareto efficient**.)

Stable matchings always exist

Theorem (Gale and Shapley 1962; RS Theorem 2.8)

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

- Gale and Shapley propose the (student-proposing) **deferred acceptance algorithm**:
- Given preferences of students and colleges, conduct the following algorithm:

Step 1 : (a) Each student "applies" to her first choice college.
(b) Each college keeps the most preferred applicant (if s/he is acceptable) and rejects all other students.

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 student (if any) held at Step (t-1), keeps the most preferred acceptable student from the combined set of students, and rejects all other students.

Termination : Stops when no more applications are made. Termination happens in finite time.

Example of DA algorithm

- Let $S = \{s_1, s_2, s_3\}$, $C = \{c_1, c_2\}$, and preferences given by

$$\succ_{s_1} : c_1, c_2,$$
$$\succ_{s_2} : c_1,$$
$$\succ_{s_3} : c_2, c_1,$$
$$\succ_{c_1} : s_3, s_2, s_1,$$
$$\succ_{c_2} : s_1, s_3.$$

- Follow steps of the DA (deferred acceptance) algorithm; I recommend you do it with a piece of paper.
- The resulting matching $\mu = \{(s_1, c_2), (s_2, \emptyset), (s_3, c_1)\}$ is stable (verify it!).

Proof of Theorem (A stable matching always exists)

The proof is very simple (but extremely elegant!!).

1. The resulting matching μ of DA is not blocked by an individual because at each step of the algorithm, no student applies to an unacceptable college and no college holds application of an unacceptable student.
2. μ is not blocked by any pair because:
   - Suppose $c \succ_s \mu(s)$ for some s and c (and s is acceptable to c).
   - This means that s applied to c and was rejected by c at some step of DA.
   - Since c's tentative match only improves as the algorithm proceeds, the match $\mu(c)$ at the end of DA is still better for c than s.
   - So c is not interested in blocking μ with s, i.e., $\mu(c) \succ_c s$.

Mechanisms in real markets

1. Stability is theoretically appealing, but does it matter in real life?
2. Roth (1984) showed that the NIMP algorithm is equivalent to a (hospital-proposing) DA algorithm, so NIMP produces a stable matching.
3. Roth (1991) studied British medical match, where different regions use different matching mechanisms. He found that stable mechanisms are successfully used (and is still in use) but most unstable mechanisms were abandoned after a short period of time.
4. In school choice, stability means "no justified envy": no student is placed in a less preferred school to another school where a student with lower priority is assigned. NYC and Boston recently adopted DA in order to, among other things, to eliminate such unfair assignment.

Mechanisms in real markets

Market	Stable	Still in use
NRMP	yes	yes (new design 98-)
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
Medical Specialties	yes	yes (1/30 no)
Canadian Lawyers	yes	yes
Dental Residencies	yes	yes (2/7 no)
Reform rabbis	yes	yes
NYC highschool	yes	yes
Boston highschool	yes	yes
Japanese Resident Matching	yes	yes

Student/College-optimal stable matchings

Theorem (Gale and Shapley 1962; RS Theorem 2.12)

There exists a **student-optimal stable matching**, that is, a stable matching that every student weakly prefers to any stable matching. The result of the student-proposing DA algorithm is the student-optimal stable matching.

Moreover, the student-optimal stable matching is college-pessimal, that is, every college weakly disprefers it to any stable matching, and vice versa (Theorem 2.13 of RS. Try to prove it yourself as this is a good exercise!)

Similarly, college-proposing DA algorithm results in the college-optimal stable matching.

- The Theorem says that different stable matchings may benefit different market participants. In particular, each version of DA favors one side of the market at the expense of the other side.
- This point was part of policy debate in NRMP in the 1990s. Recall that previous NIMP algorithm was hospital-proposing. Some medical students argued that the system favors hospitals at the expense of students and called for reconsideration of the mechanism.
- We will come back to this point in a future lecture and discuss how important this is in the context of NRMP medical match.

Proof of Theorem

Terminology: c is **achievable** for s if there is some stable matching μ such that $\mu(s) = c$. It suffices to show that no student is rejected by an achievable college in any step of DA.

For contradiction, suppose a student is rejected by an achievable college. Consider the first step in which a student (call her s) is rejected by an achievable college (call it c) (let μ be a stable matching where $\mu(s) = c$.) Then c kept some other student s' at this step, so (i) $s' \succ_c s$. Since this is the first step of DA where a student is rejected by an achievable college, (ii) $c \succ_{s'} \mu(s')$. By (i) and (ii), (s', c) blocks μ, contradicting stability of μ.

The "Rural Hospital Theorem" (RS Theorem 2.22)

Rural Hospital Theorem (RS Theorem 2.22)

The set of students and colleges that are unmatched is the same for all stable matchings.

- One motivation is the allocation of residents in rural hospitals. Hospitals in rural areas cannot fill positions for residents, and some people argue that the matching mechanisms should be changed so that more doctors end up in rural hospitals.
- But the theorem says that it is impossible as long as stable matchings are implemented.
- Also, if some students were matched in some stable matching and not in others, the latter may be unfair to him/her. The theorem says that there is no need to worry.
- In some markets, not all assumptions hold exactly, so the theorem does not hold exactly. Then it is important to know if the theorem holds approximately. I will come back to this topic in the context of NRMP in 1990.

Proof of Rural Hospital Theorem

- Let μ^S be the student-optimal stable matching and μ be an arbitrary stable matching.
- Since μ^S is student-optimal, all the students that are matched in μ are matched in μ^S (why?).
- Since μ^S is college-pessimal, all the colleges that are matched in μ^S are matched in μ (why?).
- But for any given matching, the number of matched students and colleges are the same to each other (why?).
- So the same set of students and colleges are matched in μ^S and μ (exercise: complete the argument).

Strategic behavior (RS Chapter 4)

- We have learned properties of stable matching, given information about preferences of market participants.
- But in reality, preferences are private information, so the clearinghouse should ask participants.
- Do people have incentives to tell the truth?

Strategic behavior: terminology

- A **mechanism** is a rule that produces a matching for any reported preference.
- DA is an example of a mechanism.
- A mechanism is **strategy-proof** if telling the true preferences is a **(weakly) dominant strategy** (that is, a best action no matter what others do) for everyone.

DA is not strategy-proof

- Let $S = \{s_1, s_2\}$, $C = \{c_1, c_2\}$ and

$$\succ_{s_1}: c_1, c_2,$$
$$\succ_{s_2}: c_2, c_1,$$
$$\succ_{c_1}: s_2, s_1$$
$$\succ_{c_2}: s_1, s_2.$$

- When everyone reports true preferences, DA produces $\mu = \{(s_1, c_1), (s_2, c_2)\}$.
- When c_1 reports $\succ'_{c_1}$: s_2, then DA produces $\mu' = \{(s_1, c_2), (s_2, c_1)\}$, which c_1 prefers to $\mu(c_1) = s_1$.
- So DA is not strategy-proof.

Impossibility Theorem

- DA is not strategy-proof, so people may have incentives to manipulate the mechanism.
- Unfortunately, we cannot overcome the difficulty by finding another mechanism.

Impossibility Theorem (Roth 1982; RS Theorem 4.4)

There is no stable mechanism (mechanism that produces a stable matching for all reported preferences) that is strategy-proof.

Proof

- Proof is a modification of the previous slide (available in RS page 88)
- Let $S = \{s_1, s_2\}, C = \{c_1, c_2\}$ and

$$\succ_{s_1} : c_1, c_2,$$
$$\succ_{s_2} : c_2, c_1,$$
$$\succ_{c_1} : s_2, s_1$$
$$\succ_{c_2} : s_1, s_2.$$

Thoughts on methodology

- To show impossibility, it suffices to find a particular example.
- As before, it is still important to study whether manipulation is *likely* under stable mechanisms in applications. This will be the subject in a future lecture.

DA is strategy-proof for one side

Theorem (Dubins and Freedman 1981, Roth 1982; RS Thm 4.7)

The student-proposing DA is strategy-proof for students. That is, telling the truth is a dominant strategy for every student.

- Actually it is **group strategy-proof** for students. That is, even a group of students cannot tell a lie together and make every member of the group strictly better off. See Hatfield and Kojima (2008, 2010) for general results if interested.
- Proof is skipped. Intuition: students are not punished when applying to preferred colleges (this is in a contrast with the "Boston mechanism" (aka "immediate acceptance mechanism"). A proof can be found in Theorems 10 and 11 of Hatfield and Milgrom (the model is more general and notation is different).

Many-to one matching (RS Chapter 5)

- Advance the theory to account for colleges with multiple positions.
- Everything is the same as before (Finite sets S of students and C of colleges etc) except each college c has q_c positions to fill.
- Matching μ is a correspondence from $S \cup C$ to $S \cup C \cup \{\emptyset\}$ such that
  1. $\mu(s) \in C \cup \{\emptyset\}$,
  2. $\mu(c) \subseteq S$ (each college is matched to a group of students), and
  3. $\mu(s) = c \iff s \in \mu(c)$, for every student $s \in S$ and college $c \in C$.
  4. Stability of a matching is defined similarly.

Stable matchings always exist in many-to-one matching

Theorem (Gale and Shapley 1962; RS Lemma 5.6)

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

- One easy proof: think of a college c as q_c different colleges with one position each. Then, the theorem for one-to-one matching applies.
- Or we could directly generalize the (student-proposing) DA:

Step 1 : (a) Each student "applies" to her first choice college.
(b) Each college keeps the most preferred applicants up to its quota (if s/he is acceptable) and rejects all other students.

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 student (if any) held at Step (t-1), keeps 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. Termination happens in finite time.

- Proof that DA results in a stable matching is essentially the same (good exercise!)

Many properties carry over to many-to-one matching

- Because we can think of each college c as q_c different colleges with one position, many theories of one-to-one matching carry over to many-to-one matching (so one-to-one matching theory was useful after all!). Examples:
  1. Student/college-proposing DA result in the student/college-optimal stable matchings.
  2. Rural hospital theorem: all colleges fill the same number of positions across stable matchings. Any student unmatched in any one stable matching is unmatched in all stable matchings (see also RS Theorem 5.12).

Some properties fail in many-to-one matching

- Not all properties carry over to many-to-one matching, especially strategic properties.
  1. No stable mechanism is strategy-proof for colleges (RS; Theorem 5.14). In particular, even college-proposing DA is not strategy-proof for colleges (intuition: a college is like a coalition of players in terms of strategies).
  2. On the contrary, student-proposing DA is still strategy-proof for students (why?).
  3. Colleges may benefit just by misreporting capacities. Sonmez (1997) shows that no stable mechanism is immune misreporting capacities.
  4. In one-to-one matching, DA cannot be manipulated by an agent if and only if there is a unique stable partner. The statement is false in many-to-one matching.

Married Couples (RS section 5.4.3)

- There are many married couples in the medical match (1,000 out of 20,000 in NRMP, 1990s; 30-40 out of 3,000 in psychologist match, 2000s.), and they usually want to work in the same city.
- DA fails to accommodate couples: it may assign the husband to Boston, the wife to LA, for example.
- Participation of medical students in NIMP dropped in 1970s, especially among couples.
- NIMP allowed couples to submit preferences over pairs of hospitals, and participation recovered.

There may be no stable matching with couples

- There are $C = \{c_1, c_2\}$ and one single student s and one couple (m, w)

$$\succ_s : c_1, c_2,$$
$$\succ_{(m,w)} : (c_1, c_2),$$
$$\succ_{c_1} : m, s,$$
$$\succ_{c_2} : s, w.$$

- There is no stable matching (exercise).
- So, the problem is "impossible to solve" in a sense. Then, what should we do?

Thoughts on methodology again

- A traditional way to overcome impossibility results is to find conditions on preferences under which we can say something (existence of stable matchings when couples exist, for example).
- Conditions we need is often very restrictive: "responsive" preferences (Klaus and Klijn 2005) is violated by almost all couples in data (Kojima, Pathak and Roth 2009).
- Similarly for capacity manipulations (Konishi and Unver 2006; Kojima 2007).
- But some markets like NRMP seem to be working pretty well (if one uses nice mechanisms). What makes real-life markets overcome all these impossibilities?

Summary

- Stability is important for matching in labor markets.
- Theoretically,
  1. DA produces a stable matching if the market is simple (no couples etc).
  2. Depending on which DA to use (student or college proposing), one side benefits at the expense of the other but the set of matched colleges and students do not change.
  3. DA is not strategy-proof.
  4. With couples, stable mechanisms may not work.
- Next we look at the real market and see if these theories can (or cannot) guide design of the market institution.

Design of Matching Markets

Fuhito Kojima

April 3, 2018

NRMP redesign

- We will see how basic matching theory can be used for economic design.
- But we will also see why the basic theory is not enough, and
- Other approaches are useful, and
- What kind of new theories are called for to tackle complicated issues for economic design.
- We will discuss redesign of NRMP algorithm in 1990s as a case study.

A Brief History of NRMP

- Began as a decentralized market around 1900.
- By mid 20th century, the market suffered from unraveling and congestion, causing mismatches.
- NRMP introduced a centralized matching mechanism (hospital proposing DA) in 1950s.
- Crisis in confidence in 1990s.

Crisis in Confidence in 1990s

- Some groups (such as American Medical Students Association, Public Citizen Health Research Group, Medical Student Section of the American Medical Association) advocated reconsideration of the algorithm.
- The Board of Directors of NRMP commissioned the design of a new algorithm.
- New algorithm based on student-proposing DA (but accommodating couples and other complications) was introduced from 1998.
- The algorithm is in use now.

What were the issues?

1. Does the NRMP algorithm favor hospital programs at the expense of doctors?
   $\rightarrow$ Yes, since NRMP is the hospital-proposing DA.
2. Is NRMP a "manipulable" system where students should report false preferences to get the best outcome?
   $\rightarrow$ Yes, since both doctors and hospitals may have incentives to manipulate.

NRMP "match variations"

As suggested before, NRMP has special features, called "match variations," which is not present in the simple theory. Examples are:

1. Couples,
2. Hospital programs that want to fill even number of positions,

Further problems with match variations

There are problems that happen because there are match variations.

1. Some people are unmatched because of the choice of the algorithm?
   $\rightarrow$ No such concern if no match variations are present (rural hospital theorem), but possible otherwise.
2. Does NRMP find a stable matching in the first place?
   $\rightarrow$ A stable matching exists if no match variations are present, but a stable matching may not exist otherwise.

Empirical study

1. Traditional theory points to potential problems even without match variations.
   1. hospitals are favored
   2. the system is manipulable
2. With complex reality (e.g., couples), positive theoretical results do not apply
   1. different numbers of students may be matched in different stable matchings
   2. a stable matching may not exist
3. To evaluate these concerns, empirical and numerical studies are useful. So, look at the market!
   $\rightarrow$ Roth and Peranson (1999) obtained data on NRMP such as submitted preferences, and conducted simulations.

Descriptive statistics of NRMP

	1987	1993	1994	1995	1996
APPLICANTS					
Applicants with ROLs	20071	20916	22353	22937	24749
Applicants who are Coupled	694	854	892	998	1008
PROGRAMS					
Active Programs with ROL	3170	3622	3662	3745	3758
Total Quota Before Match	19973	22737	22801	22806	22578

Difference between hospital proposing and college proposing DAs

	1987	1993	1994	1995	1996
APPLICANTS					
Number of Applicants Affected	20	16	20	14	21
Applicant Proposing Preferred	12	16	11	14	12
Program Proposing Preferred	8	0	9	0	9
New Matched	0	0	0	0	1
New Unmatched	1	0	0	0	0
PROGRAMS					
Number of Programs Affected	20	15	23	15	19
Applicant Proposing Preferred	8	0	12	1	10
Program Proposing Preferred	12	15	11	14	9
Prog. with New Position(s) Filled	0	0	2	1	1
Prog. with New Unfilled Positions	1	0	2	0	0

Magnitude of possible manipulations by students

Upper limit of the number of applicants who could benefit by truncating their lists at one above their original match point (for students, truncation is known to be "exhaustive"[1])

	1987	1993	1994	1995	1996
Program-Proposing Algorithm	12	22	13	16	11
Applicant-Proposing Algorithm	0	0	2	2	9

As expected, more applicants can benefit from list truncation under the program-proposing algorithm than under the applicant-proposing algorithm.

But both numbers are very small.

[1] Roth and Vande Vate (1991).

Magnitude of possible manipulations by hospitals

Upper limit of the number of hospital programs that could benefit by truncating their lists at one above their original match point (for hospitals, truncation is not exhaustive[2])

	1987	1993	1994	1995	1996
Program-Proposing Algorithm	15	12	15	23	14
Applicant-Proposing Algorithm	27	28	27	36	18

[2] Kojima and Pathak (2009) show that the class of "dropping strategies" is exhaustive.

Magnitude of possible manipulations by capacities

Hospitals can also misreport capacities (Sonmez 1997).[3]

Estimate of the Upper Bound of the Number of Programs That Could Improve Their Remaining Matches By Reducing Quotas

	1987	1993	1994	1995	1996
Program Proposing Algorithm	28	16	32	8	44
Applicant Proposing Algorithm	8	24	16	16	32

In fact, hospitals can manipulate *both* ranking and capacities, and it may not need to use truncation (but this was not done by Roth and Peranson).

[3] Tayfun Sonmez (1997), "Manipulation via Capacities in Two-Sided Matching Markets," Journal of Economic Theory, shows that no mechanism is immune to capacity manipulation.

Why is looking at data not sufficient?

All results above are suggestive that prediction of simple theories are approximately correct, and some of the potential problems suggested by theories may not be important.

But looking at data alone is only suggestive, and not conclusive.

We will look at two additional approaches:

1. simulation on randomly generated data, and
2. theoretical analysis.

Magnitude of conflict of interest/manipulations

Simulation on randomly generated data.

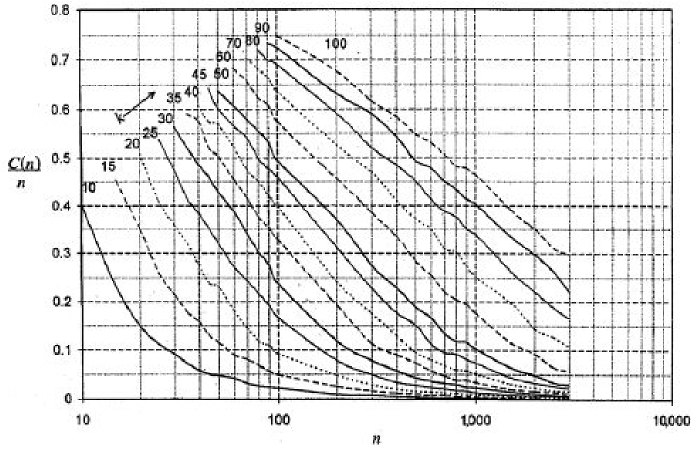
Simple model: n hospital programs, n doctors, (no couples).

Preferences are drawn independently and uniformly. Each doctor applies to k hospitals.

$C(n) =$ number of doctors matched differently at hospital-proposing and doctor-proposing DAs.

- In one-to-one markets, an agent can profitably manipulate if and only if s/he is not matched to his/her optimal stable partner $\rightarrow$ C(n) is the number of agents who can manipulate DA!

Magnitude of possible manipulations



Large size of the market is the key!

Theory: Incentives in Large Matching Markets

Based on Kojima and Pathak (2009): See also Immorlica and Mahdian (2005).

Look at an example with manipulation possibilities. Two students $\{i, j\}$ and two colleges $\{S\,(\text{Stanford}), H\,(\text{Harvard})\}$, with one seat each.

$$\begin{aligned} \succ_i &: S, H, \\ \succ_j &: H, S, \\ \succ_S &: j, i, \\ \succ_H &: i, j. \end{aligned}$$

- If everyone is truthful:
  S is matched to i, H is matched to j.
- If S declares i unacceptable ($\succ'_S : j$):
  S is matched to j, H is matched to i.

College S successfully manipulates DA, but this happens because of a very subtle "rejection chain".

Main Result

Kojima and Pathak (2009) set up a model in which student preferences are randomly drawn, and consider the probability that any one college can profitably misreport as the market becomes large (i.e., lots of students and colleges).

Theorem

For any $\varepsilon > 0$, there exists n such that no colleges can profitably misreport with more than probability ε while other colleges are reporting true preferences.

Intuition

Strategic rejection by a college causes a chain of application and rejections. Some of the rejected students may apply to the manipulating college, and the college may be made better off if these new applicants are desirable.

In a large market, there is a high probability that there will be many colleges with vacant positions. So the students who are strategically rejected (or those who are rejected by them and so on) are likely to apply to those vacant positions and be accepted. So the manipulating college is unlikely to be made better off.

Sketch of Proof (Step 1): Dropping Strategy

$(\succ'_c, q'_c)$ is a **dropping strategy** of $(\succ_c, q_c)$ if
(1) $q'_c = q_c$, and
(2) $\succ'_c$ drops some acceptable students from $\succ_c$, but does not change orders between remaining students.

Lemma

If c cannot manipulate student-proposing DA successfully by a dropping strategy, then c cannot manipulate it successfully by any strategy.

This lemma simplifies analysis by narrowing down the class of strategies to consider.

Sketch of Proof (Step 2): Rejection Chains

Given c and dropping strategy $\succ'_c$, consider **rejection chains** algorithm, an algorithm similar to student-proposing DA:
(1) First, run DA under true preferences.
(2) Then let c reject students matched to c who are unacceptable under $\succ'_c$. Each rejected student applies to next choice, just as in DA. The rest proceeds as in DA.
The rejection chain **returns to** c if some student applies to c at Step (2).

Lemma

If no rejection chains return to c, then no dropping strategies are successful manipulations for c.

Sketch of Proof (Step 3): Vanishing Market Power

Lemma (Vanishing market power)

For any $\varepsilon > 0$, if the number of colleges n in the market is sufficiently large,

$$\Pr(\textit{at least one rejection chain returns to } c) < \varepsilon$$

for any college c in the market.

Intuition: In a large market, with high probability there are many colleges with vacant positions. So the rejected students (or those who are rejected by them and so on) usually apply to those vacant positions and are accepted, ending a rejection chain.
Lemmas 1-3 show the theorem.

New Design (Roth-Peranson algorithm)

The new (and current) NRMP algorithm, called the Roth-Peranson algorithm, is based on student-proposing DA, but try to accommodate couples.

The algorithm allows couples to express preferences on pairs of hospital programs.

First run DA without couples, and then add couples one at a time.

If someone is displaced, then such an agent is allowed to apply later in the algorithm.

The basic idea is based on Roth and Vande Vate (1989) on one-to-one matching.

Aside: Some open questions

Large-market incentive results have been studied/extended by Lee (2013), Ashlagi, Kanoria, and Leshno (2014), Coles and Shorrer (2013), and Storms (2014). Any general result?

Roth and Vande Vate (1989) showed that, starting from any matching, there is a sequence of blocking pairs that leads to a stable matching in one-to-one matching without couples.

Kojima and Unver (2008) showed a similar result in many-to-many matching, when one side has substitutable preferences and the other side "responsive" preferences.

One conjecture is that the same result holds when every agent has substitutable preferences (it is known that stable matching exists when every agent has substitutable preferences, and it is essentially the weakest such condition.)

Couples

In NRMP and in other markets, couples (and other match variations) make it possible for nonexistence of stable matchings, and failure of the rural hospital theorem.

- Example: $C = \{c_1, c_2\}$, one single student s and one couple (m, w)

$$\begin{aligned} \succ_s &: c_1, c_2, \\ \succ_{(m,w)} &: (c_1, c_2), \\ \succ_{c_1} &: m, s, \\ \succ_{c_2} &: s, w. \end{aligned}$$

Also conclusions such as non-manipulability of DA in large markets are not directly applicable. Even worse, DA may not be strategy-proof even for students.

Couples

Despite theoretical difficulty, the stable matching mechanism seems to be used smoothly. Here is a quote from Roth (2008):[4]

[An] empirical observation made in the resident match data, and in the other matches ... is that, even when couples are present, it is a very rare occurrence for the set of stable matchings to be empty. ... An open question is why this is so.

[4] Roth, Alvin E. "Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions," International Journal of Game Theory, Special Issue in Honor of David Gale on his 85th birthday, 36, March, 2008, 537-569.

Matching with couples: Theory

Kojima, Pathak and Roth (2013) consider a model similar to Kojima and Pathak but assume there are a 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.

*Ashlagi et al. (2014) relax some of the assumptions for Kojima, Pathak, and Roth's (2013) result. Related papers by Biro et al. (2014), Nguyen and Vohra (2016), Che, Kim, and Kojima (2018)

Intuition for Existence

Roth-Peranson algorithm will find a stable matching if couples are not displaced by another couple or single doctors.

In a large market, there is a high probability that there will be many colleges with vacant positions. So couples and singles are unlikely to apply and displace a couple in a hospital. So the algorithm is likely to terminate, producing a stable matching.

Summary

DA is not strategy-proof, and there are also some other potential problems. Why is the mechanism adopted in applications?

We took NRMP as case study and studied how big such problems are.

Numerical studies based on real data and simulation suggest that they are not large problems.

Inspired by observations in such markets, some new theories are developed to evaluate performance of stable mechanisms.