Networks II: Market Design—Lecture 17 Matching Markets with Non-transferable Utilities: Two-sided preferences

ARPITA GHOSH

Dept. Of Information Science,
Cornell University

Logistics

- Preliminary report for project: Due **Thursday**, April 11 by noon; on CMS
  - Length: 2 pages (or less, but not more)
  - Weight: 7%
  - Usual late policy applies
  - Remember to make one submission per group, *with all members included*!

- Reminder, for post-spring break: Midterm April 18th, in-class
  - Recall: CMS writeups (Course Information, ...)
  - **No alternative exam: Make sure to be here!**

Preliminary report guidelines

Recap: Project content

- Topic: *Anything* related to networked behavior
  - Can pick topic of interest from Networks I, II, or out-of-class
  - **Must** make connection to networked behavior: See Lecture 1
  - Utilizing what you're learning in this class: Value of abstract model
- Ideal project: Depth on *at least* one dimension, touches on other two
  - Identification and abstract modeling
  - Analysis
  - Design

Recap: Evaluation criteria for project

- Significance: How is your project related to networked behavior and how do you benefit from an abstract model?
- Clarity: What exactly are you doing, and what are you doing about it?
- Coherence: How do the various 'what are you doing's connect to each other?
  - Especially important for literature survey
- Value-add: Original contribution (insight, commentary, modeling, problem-solving, . . .)

Preliminary report: Guidelines

- Purpose: Preliminary report provides (incentive for, and) evidence of progress towards completion
- 'What should my 2 page-report consist of?'
- Preliminary report: (i) Reports progress so far (ii) Specifies what will be done in remainder
  - **Must** identify problem being addressed; make connection to networked behavior
  - Must demonstrate partial progress on (at least) depth dimension
  - Must contain clear list of steps remaining to complete project

Preliminary report: **Example** guidelines

- (i) Identify a real-world networks setting; provide possible formal models/abstractions for it *and identify what the purpose of the model is*; identify assumptions and issues (depth on identification and modeling) (ii) Summarize literature to suggest analysis/design solutions
- (i) Identify (simple but well-motivated) extension on existing model (ii) Provide (progress towards) analysis or design solution
- Literature surveys: (i) Identify area being surveyed (ii) Insights from at least 2 papers (iii) Summary of remaining papers to be covered, and why

Preliminary report: Grading criteria

- 'How will my 2 page-report be graded?'
- A preliminary report: (i) Reports progress so far (5pts) (ii) Specifies what will be done in remainder (2pts)
- Every report must:
  - Clearly identify problem being addressed (2pts): (i) Motivate setting (1pt) (ii) Identify a clear question to be addressed (1pt)
  - Make connection to networked behavior (1pt)
  - Demonstrate partial progress (at least) on depth dimension (2pts)
  - Contain clear list of steps remaining to complete project (2pts)

Recap: Last time

Comparing matchings:

- Definition: $\mu_1 \succeq_M \mu_2$ if **all** men $m \in M$ (weakly) prefer partner in μ_1 to partner in μ_2 ($\mu_1 \succeq_W \mu_2$ defined similarly)
- Main result: *Stable matching μ_M^* returned by man-proposing DA is man-optimal amongst all stable matchings* (and similarly for women)
- Why you should be gaping:
  - Men agree on best stable matching, *despite* competing with each other
  - This matching is returned by man-proposing DA where men's partners get steadily worse
  - Set of pairs $(m, w^*(m))$ where $w^*(m)$ is m's favorite achievable partner *forms a matching*, and this matching is stable!

Man and woman-optimal stable matchings

Theorem

For any two **stable** *matchings* μ *and* μ', $\mu \succeq_M \mu'$ *if and only if* $\mu' \succeq_W \mu$.

Why is this true?

- Assume that $\mu \succeq_M \mu'$: we'll show $\mu' \succeq_W \mu$
- Suppose not: There must be woman $w \in W$ such that $\mu(w) \succ_w \mu'(w)$; let $m = \mu(w)$
- By assumption, $w = \mu(m) \succ_m \mu'(m)$. Also, $m \succ_w \mu'(w)$
- But then (m, w) is a blocking pair for μ'
- So μ' cannot have been stable: Contradiction!
    - Note strict preferences!

Women-pessimal matchings

Corollary

The man-optimal stable matching is women-pessimal.
Similarly, the woman-optimal stable matching is man-pessimal.

- μ^*_M: Outcome of man-proposing DA
- μ^*_M is man-optimal stable matching: $\mu^*_M \succeq_M \mu$ for any stable matching μ
- For any stable matching μ: $\mu^*_M \succeq_M \mu \Rightarrow \mu^*_M \preceq_W \mu$ (Previous theorem)
- This holds for all stable matchings μ: μ^*_M is woman-pessimal!
- Identical argument for second statement

Another example

- Consider market with preferences:

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_2	w_2	m_2	m_1	m_2
w_3	w_3	w_3	m_3	m_2	m_3

- What is outcome of DA with men proposing?
  - $\mu_M = (m_1 - w_1, m_2 - w_3, m_3 - w_2)$
- Consider matching $\mu = (m_1 - w_2, m_2 - w_3, m_3 - w_1)$
- Which is true?
  [A] $\mu_M \succeq_M \mu$ [B] $\mu \succeq_M \mu_M$ [C] μ_M and μ are not comparable
  - What's going on?
  - Is matching μ stable?
  - No (blocked by pair (m_2, w_1)): Result only compares *stable matchings*!

Check-in

- Understanding results so far: ([A] True [B] False)
  - $\mu^*_W \succeq_W \mu$ for all matchings $\mu \neq \mu^*_W$
    - μ^*_W: Outcome of woman-proposing DA
  - For any two stable matchings, either $\mu_1 \succeq_M \mu_2$ or $\mu_2 \succeq_M \mu_1$
  - For any two stable matchings, either $\mu_1 \succeq_M \mu_2$ or $\mu_1 \succeq_W \mu_2$
  - Given any stable matching μ, there exists some matching that all men weakly prefer in comparison with μ

- Coming up next:
  - Unmatched agents in stable matchings
  - Incentives!

Unmatched agents

- With either unequal number of men and women, or incomplete lists, someone might go unmatched
- How does the set of unmatched agents change across stable matchings?

An example

- 3 men (m_1, m_2, m_3) and 3 women (w_1, w_2, w_3)
- Preferences:

m_1	m_2	m_3	w_1	w_2	w_3
w_1	w_3	w_3	m_2	m_2	m_1
w_2	w_1	w_1	m_1		m_3
w_3	w_2				m_2

- What is outcome of DA with men proposing?
- What is outcome of DA with women proposing?

Another example

- 4 men (m_1, m_2, m_3, m_4) and 2 women (w_1, w_2)
- Preferences:

m_1	m_2	m_3	m_4	w_1	w_2
w_1	w_1	w_2	w_2	m_3	m_1
w_2	w_2	w_1		m_4	m_3
				m_1	m_2
				m_2	m_4

- What is outcome of DA with men proposing?
  - $(m_1 - w_1, m_3 - w_2)$
- What is outcome of DA with women proposing?
  - $(m_1 - w_2, m_3 - w_1)$

Is there a general principle here?

Theorem

The set of men and women that are unmatched is the same in all *stable matchings. (!)*

Why is this true?

- Let μ be an arbitrary stable matching, and μ_M^* be the man-optimal stable matching (we know this exists!)
- Notation:
  - Let $M_{\mu_M^*}$, M_μ be **set** of men matched in matchings μ_M^*, μ
  - Similarly define $W_{\mu_M^*}$, W_μ
- μ_M^* is man-optimal: $M_{\mu_M^*} \supseteq M_\mu$
- μ_M^* is woman-pessimal: $W_{\mu_M^*} \subseteq W_\mu$
- $|W_{\mu_M^*}| \leq |W_\mu| = |M_\mu| \leq |M_{\mu_M^*}| (= |W_{\mu_M^*}|!)$: So all inequalities must hold with equality, implying $M_{\mu_M^*} = M_\mu$, $W_{\mu_M^*} = W_\mu$!
- Question: Why does this argument only work for *stable* matchings μ?

Where we are

- Deferred acceptance (DA) algorithm:
  - Returns a stable matching
  - Man-proposing DA returns man-optimal stable matching
  - Woman-proposing DA returns woman-optimal stable matching
- Same number of matched agents in all stable matchings
- What's next?

Strategic behavior

- DA algorithm: All properties wrt *reported preferences*
- Do agents report their true preferences?
- Recall:
  - *Mechanism* is a function that returns a matching for any input (set of agents and reported preferences)
    - Again: Mechanism is algorithm with input from strategic agents (Eg: Deferred acceptance (DA) algorithm)
  - A mechanism is **strategyproof** if reporting true preferences is a **dominant strategy** (that is, a best action no matter what others do) for all agents
- How does DA algorithm do as a *mechanism*?

Flashback: Mechanisms in one-sided markets

- Serial dictatorship mechanism:
  - Theorem: Pareto-efficiency wrt *reported preferences*
  - Do agents report their true preferences?
  - Theorem: Serial dictatorship mechanism is strategyproof
- TTC algorithm:
  - Theorem: TTC algorithm returns a core matching
  - Theorem: TTC algorithm is strategyproof (Roth'82)
- What about the DA algorithm?

Incentives in DA

- Two agents on each side $\{m_1, m_2\}, \{w_1, w_2\}$
- Preferences:

$$\succ_{m_1} : w_1, w_2,$$
$$\succ_{m_2} : w_2, w_1,$$
$$\succ_{w_1} : m_2, m_1$$
$$\succ_{w_2} : m_1, m_2.$$

- Matching produced by (man-proposing) DA with true preferences μ: $(m_1 - w_1, m_2 - w_2)$
- Suppose w_1 reports $\succ'_{w_1}$: m_2
- DA produces μ': $(m_1 - w_2,\ m_2 - w_1)$, which w_1 prefers to $\mu(w_1) = m_1$!

- Example shows: DA is not strategy-proof!
  - 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.*

Proving the result

- Same market as before: Preferences P are
  $\succ_{m_1}: w_1, w_2; \quad \succ_{m_2}: w_2, w_1; \quad \succ_{w_1}: m_2, m_1; \quad \succ_{w_2}: m_1, m_2$
- Stable matchings (for true preferences P):
  $\mu_1 = (m_1 - w_1, m_2 - w_2); \quad \mu_2 = (m_1 - w_2, m_2 - w_1)$
- *Any* stable matching mechanism M *must* return μ_1 or μ_2 when P is reported
  - If M returns μ_1: If w_1 reports preference m_2, M must return $\mu' = \{(m_1, w_2), (m_2, w_1)\}$, so w_1 benefits from lying
  - If M returns μ_2: m_1 can profitably lie!
- So M cannot be simultaneously strategyproof and a stable matching mechanism