Networks II: Market Design—Lecture 15 Matching markets with two-sided preferences: The Deferred Acceptance Algorithm

ARPITA GHOSH

Dept. Of Information Science,
Cornell University

What we saw last time

- How does stability relate to Pareto-efficiency, core?
  - Stable matchings are Pareto-efficient, but not all Pareto-efficient matchings are stable
  - Every stable matching is in the core, and every core matching is stable
  - Comparing two definitions: How to, and answers

Immediate questions

- How does stability relate to Pareto-efficiency, core in two-sided markets?
  - Stable matchings are Pareto-efficient
  - But not all Pareto-efficient matchings are stable
  - Every stable matching is in the core, and every core matching is stable
- Does a stable matching always *exist*? If yes, how can we find one?
- What are the properties of stable matchings?
- Stability is a pretty abstraction, but does it matter in reality?

Finding a stable matching

- A natural idea: Start with some arbitrary matching μ
- 'Fix' if not stable, by matching a blocking pair: Does this work?
- μ_0: $m_1 - w_1, m_2 - w_2, m_3 - w_3$ (preferences as before)

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

  - Start with blocking pair $m_2 - w_1$, to get new matching μ_1: $m_1 - w_2, m_2 - w_1, m_3 - w_3$
  - Is this stable?
    - A Yes
    - B No

Finding a stable matching

- Try this again: μ_0 is $m_1 - w_2, m_2 - w_3, m_3 - w_1$ (preferences as before)

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

  - Blocking pair (m_3, w_2): New matching $m_1 - w_1, m_2 - w_3, m_3 - w_2$
  - Blocking pair (m_2, w_1): New matching $m_1 - w_3, m_2 - w_1, m_3 - w_2$
  - Is this matching stable?
    - A Yes
    - B No

- Yet another example: Suppose 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

- Start with μ_0: $m_1 - w_1, m_2 - w_2, m_3 - w_3$
  - Blocked by (m_1, w_2): Switch to $m_1 - w_2, m_2 - w_1, m_3 - w_3$
  - Blocked by (m_3, w_2): Switch to $m_1 - w_3, m_2 - w_1, m_3 - w_2$
  - Blocked by (m_3, w_1): Switch to $m_1 - w_3, m_2 - w_2, m_3 - w_1$
  - Blocked by (m_1, w_1): Switch to $m_1 - w_1, m_2 - w_2, m_3 - w_3$
- Iterative unblocking *cycles* back to original matching μ_0!

A natural question

- Example due to Knuth
- Open problem (Knuth, 1976): Is there an initial matching, and sequence of blocking pairs, such that iterative 'correction' converges to a stable matching?
- (Eventually) resolved by Roth & Vande Vate, in 1990(!)
    - Algorithm to find a stable matching by iteratively eliminating blocking pairs
    - Analysis not so simple: But turns out to be central in more advanced applications

Finding a stable matching

- Recall: "Immediate questions"
  - Do stable matchings always exist? How can we find them?
- Natural algorithm: Start with a matching; fix if not stable
  - 'Greedy' procedure does not work: Can cycle back to original matching (Knuth 1976)
  - Iterative unblocking conjecture resolved in 1990

Existence

Do stable matchings always exist? And how can we find them?

Theorem (Gale & Shapley, 1962)

For any instance of the stable marriage problem, there always exists a stable matching.

Proof by *construction!*: Simultaneously answers both questions

- Such a matching always exists
- Algorithm for constructing a stable matching

The Deferred Acceptance Algorithm (Gale-Shapley, 1962)

- Initialize: All men and women are free
  (Assume strict, complete preferences)
- While there is a man m who is free and hasn't proposed to every woman
  - Choose such a man m: Let w be m's most-preferred choice that m has not yet made a proposal to
  - m proposes to w
  - If w is free, then (m, w) become *engaged*
  - Else, if w is currently engaged to m':
    If w prefers m' to m, then m remains free
    If w prefers m to m', then (m, w) become engaged; m' becomes free (ha!)

The Deferred Acceptance Algorithm (Gale-Shapley, 1962)

- A 'centralized' version:
- In each *round*:
  - Each ‘free’ man m proposes to his top-ranked woman w who has not rejected him yet
  - Each woman chooses amongst her current ‘provisional’ partner, and the best proposal from the current round
  - Woman is engaged to this chosen man at the end of the round, and rejected men become free
- When no more proposals can be made, all engagements are finalized
- Return resulting matching

Two notes:

- Does the order in which men propose matter in the decentralized version?
  - 'Matter': Is outcome of algorithm independent of order of proposals?
- Do centralized and decentralized versions of algorithm produce same outcome?

- No, and yes: Verify for yourself why!

The Deferred Acceptance Algorithm (Gale-Shapley, 1962)

(Centralized version)

- In each *round*:
  - Each 'free' man m proposes to his top-ranked woman w who has not rejected him yet
  - Each woman chooses amongst her current 'provisional' partner, and the best proposal from the current round
  - Woman is engaged to this chosen man at the end of the round, and rejected men become free
- When no more proposals can be made, all engagements are finalized
- Return resulting matching

The deferred acceptance algorithm: Does it work?

To show this algorithm 'works', we need to show:

- Convergence: Does the algorithm always terminate?
  - Yes: Each man proposes to each woman at most once
    —Recall: "Each 'free' man m proposes to his top-ranked woman w who has not rejected him yet"
  - Finite number of men, women
- Correctness: Does it return a stable matching?

Deferred acceptance: Example

- 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_3	w_2	m_3	m_1	m_3
w_3	w_2	w_3	m_2	m_2	m_2

- How does the deferred acceptance algorithm work on this example?

Consider *centralized version* of Deferred Acceptance algorithm
(Recall: All agents have (strict) *complete* preferences)

- In each round of the algorithm, ([A] True [B] False)
  - Each woman receives a proposal
  - Each man makes a proposal
- DA terminates in exactly n rounds, if there are n men and n women in the market
- After the last-but-one round of the algorithm,
  - At least one man is free
  - Every woman has received at least one proposal, and therefore no woman is un-engaged
- For any market size n, there exists a preference profile for which DA terminates in one round

Consider *centralized version* of Deferred Acceptance algorithm (Recall: All agents have (strict) *complete* preferences)

- Suppose m prefers $w_7 \succ w_3 \succ w_1 \succ \dots$. If DA (Deferred Acceptance) matches m to w_3, then:
  - A m cannot have proposed to w_1 (in any preference profile)
  - B m must have proposed to w_1 (in every preference profile)
  - C m may have proposed to w_1 (in some preference profile)
- What if w_1 is replaced by w_7 in this question?
- Suppose w prefers $m_7 \succ m_3 \succ m_1$. If the DA algorithm matches w to m_3, then:
  - A m_1 cannot have proposed to w (in any preference profile)
  - B m_1 must have proposed to w (in every preference profile)
  - C m_1 may have proposed to w ('in some' preference profile)
- What if m_1 is replaced by m_7 in the questions above?

Analyzing correctness: Some facts

For every man m,

1. The sequence of women to whom m proposes gets worse and worse (in terms of $\succ_m$)
   —Each 'free' man m proposes to his top-ranked woman w who has not rejected him yet

For every woman w,

2. w is always engaged after the point at which she receives her first proposal
   —Each woman chooses amongst her current 'provisional' partner, and the best proposal from the current round
3. w has an *improving* sequence of partners (in terms of her preference list $\succ_w$)

Correctness: Will it return a stable outcome?

- Suppose matching μ returned by algorithm is not stable:
  - Unstable: Blocked by pair (m, w)
  - m prefers w to $w' = \mu(m)$ and w prefers m to $m' = \mu(w)$
- m's last proposal in DA was to w' (Why?)
- m prefers w to w': m should have proposed to w before w' in DA (Fact [1])
- So m was rejected by w, which means w should be matched to a man she likes better than m
  - Sequence of partners of w only improves, by Fact [3]
- So (m, w) is not a blocking pair: μ must be stable

Existence and construction of stable matchings

Theorem (Gale & Shapley, 1962)

The deferred acceptance algorithm returns a matching that is stable for any instance of the marriage model.

- DA algorithm simultaneously answered both our questions:
  - Does a stable matching always exist?
  - How can we find one?

Extending the basic marriage model

Two easy extensions of the model:

(i) Different numbers of men and women: $|M| \neq |W|$

(ii) *Incomplete* lists:

- Man m may prefer to remain unmatched than be matched to w (similarly for women)
- Formally: A man's preference is now a ranked ordering of $W \cup m$
- w 'not on' m's preference list: $m \succ_m w$
  - Man m prefers staying single to being matched to w
- (Equivalent definitions for $w \in W$)

Matchings with incomplete lists

- Matching μ: Function from $M \cup W$ to $M \cup W$ such that
  - $\mu(m) \in W \cup m$, and $\mu(w) \in M \cup w$;
  - $\mu(m) = w \iff \mu(w) = m$, for every man $m \in M$ and woman $w \in W$
- Stable matching with incomplete preferences:
  - Matching μ is **blocked by an individual** a if $\mu(a)$ is unacceptable to a
  - **Individual rationality**: μ is not blocked by any $a \in M \cup W$
  - A matching is stable if it is (i) individually rational *and* (ii) contains no blocking pairs

- When there are different numbers of men and women, and preferences can be *incomplete*,
  - Does a stable matching always exist?
  - How can we find one?

The Deferred Acceptance Algorithm

Deferred Acceptance algorithm with incomplete lists:

- In each round:
  - Each 'free' man m proposes to his top-ranked **(acceptable)** woman w who has not rejected him yet
  - Each woman chooses amongst her current 'provisional' partner, and the best **(acceptable)** proposal from the current round
  - Woman is engaged to this chosen man **(if such exists)** at the end of the round; rejected men become free
- When no more proposals can be made, all engagements are finalized
- Return resulting matching

Analysis of Deferred Acceptance: I

- Convergence: Does the algorithm always terminate?
  - Yes: Each man proposes to each woman at most once
  - Finite number of man-woman pairs
- Observations 1-3 continue to hold, rephrased slightly:
  1. For every man m, the sequence of women to whom m proposes gets worse and worse (in terms of $\succ_m$)

  2,3. For every woman w,
     - Once engaged, w continues to remain engaged through the algorithm
     - w has an *improving* sequence of partners (in terms of her preference list $\succ_w$)

  4. No man proposes to an unacceptable woman; no woman gets engaged to an unacceptable man

Correctness: Will it return a stable outcome?

- Suppose μ is not stable: Blocked by (m, w)
    - m prefers w to $\mu(m)$ and w prefers m to $\mu(w)$
    - No proposals made to, or accepted from, unacceptable partner: No blocking individuals
- m must have proposed to w: He finds her acceptable, and prefers her to $\mu(m)$
- So w rejected m: Either m is not acceptable to w, or w received a proposal from a man m' that she prefers to m
    - True [A] or false [B]: m' is w's partner in DA's outcome $(m' = \mu(w))$
- In either case (m, w) is not a blocking pair: So μ is stable

Stable matchings in a market

- Preferences:

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

- Deferred-acceptance gives matching $(m_1 - w_1, m_2 - w_2, m_3 - w_3)$
- Can you identify another stable matching in this market?
  - Women proposing also gives stable matching: $(m_1 - w_2, m_2 - w_3, m_3 - w_1)$
- Is there any other stable matching in this market?
  - Yes: Matching $(m_1 - w_3, m_2 - w_1, m_3 - w_2)$ is stable as well!
- Deferred acceptance algorithm returns **one** stable matching: Can be other stable matchings too!

So far

- Deferred acceptance (DA) algorithm returns stable matching
  - Argument assuming (strict) complete preferences, $|M| = |W|$
  - Extend model to $|M| \neq |W|$; (strict) incomplete preferences
  - Incomplete preferences: Stability means no blocking pairs **and** individual rationality
- Appreciating the result:
  - Analog of TTC algorithm for two-sided markets
  - Natural greedy algorithm does not work
  - Stable matchings need not always exist in all markets! (HW4)
- Deferred acceptance produces **a** stable matching: Other stable matchings may exist!
- Next: Comparing matchings