Mechanismen ohne Geld

Algorithmische Spieltheorie

Sommer 2017

Mehrheitsregel mit zwei Kandidaten

Direkte Kanzlerwahl: (M)erkel, (S)chulz

Wähler	Präferenzordnung
1	M S
2	S M
3	M S

Wähler	Genannte Ordnung
1	M S
2	M S
3	M S

Resultat der Mehrheitsregel: M, S

Mehrheitsregel für zwei Kandidaten setzt viele gute Eigenschaften um:

- Repräsentiert die Mehrheit der Präferenzen
- Jeder Kandidat in der Position, in der er/sie am meisten genannt wird
- Strategisches Wählen ist nicht profitabel:
  Wähler mit Präferenz wie die Mehrheit ändert sein Votum: Kein besseres Ergebnis. Wähler mit Präferenz entgegen der Mehrheit kann Ergebnis nicht ändern.

Drei Kandidaten

Direkte Kanzlerwahl: (M)erkel, (S)chulz, (N)ocheiner

Wähler	Präferenzordnung
1	M S N
2	S N M
3	N M S

Mehrheitswahl ergibt einen Kreis:
Jeweils 2 Wähler mögen M lieber als S, S lieber als N, und N lieber als M ...

Diese Instanz zeigt, dass die kollektive Präferenz widersprüchlich sein kann (kreisend, nicht transitiv) obwohl jede individuelle Präferenz wohl-definiert ist.

Das Beispiel wird Condorcet-Paradox genannt und wurde vom Marquis de Condorcet gegen 1785 entdeckt.

Pluralitätswahl

Wir betrachten die Pluralitätsregel (plurality rule), bei der jeder Kandidat auf die Position in der Rangliste kommt, auf der er am häufigsten bei den Wählern auftritt. Wir lösen Ties bzgl. alphabetischer Ordnung.

Wähler	Präferenzordnung
1	M S N
2	S N M
3	N M S

Pluralität: M, N, S

Wähler	Genannte Ordnung
1	M S N
2	S N M
3	N S M

Pluralität: N, S, M

Strategisches Wählen ist profitabel für den dritten Wähler!

Wie können wir strategisches Wählen vermeiden? Trivialerweise können wir einen Wähler als Diktator bestimmen, der das Ergebnis diktiert durch seine genannte Präferenz. Aber bekommen wir das auch anders hin?

Definitionen

- Menge A von Kandidaten (oder Ergebnissen, Alternativen)
- Menge $\mathcal{N}$ von n Wählern (oder Spielern)
- Menge L der möglichen Präferenzen (totale Ordnungen von A)
- Jeder Wähler i hat eine Präferenz (oder Präferenzordnung) $\succ_i \in L$ über die Kandidaten A

- Eine soziale Nutzenfunktion ist eine Funktion $F : L^n \rightarrow L$.
- Eine soziale Auswahlfunktion ist eine Funktion $f : L^n \rightarrow A$.

Eine soziale Auswahlfunktion gibt nur einen einzigen Sieger aus, eine soziale Nutzenfunktion gibt eine vollständige Rangliste aller Kandidaten aus.

Eigenschaften von sozialen Nutzenfunktionen

- Einstimmigkeit: Für jedes $\succ \in L$ gilt $F(\succ, \ldots, \succ) = \succ$.
- Wähler i ist ein Diktator in einer sozialen Nutzenfunktion wenn für alle $\succ_1, \ldots, \succ_n \in L$ gilt $F(\succ_1, \ldots, \succ_n) = \succ_i$. Dann wird F eine Diktatur genannt.
- Unabhängigkeit von irrelevanten Alternativen (IIA): Die soziale Präferenz zwischen je zwei Kandidaten a und b hängt nur von den Präferenzen der Wähler bzgl. a und b ab.

  Formal: Für jede $a, b \in A$ und jede $\succ_1, \ldots, \succ_n, \succ'_1, \ldots, \succ'_n \in L$, sei $\succ = F(\succ_1, \ldots, \succ_n)$ und $\succ' = F(\succ'_1, \ldots, \succ'_n)$. Wenn $a \succ_i b \Leftrightarrow a \succ'_i b$ für alle i, dann gilt $a \succ b \Leftrightarrow a \succ' b$.

Unabhängigkeit von irrelevanten Alternativen (IIA)

Pluralität verletzt IIA!

Wähler	Präferenzordnung
1	M S N
2	S N M
3	N M S

Pluralität: M, N, S

Wähler	Genannte Ordnung
1	M S N
2	S N M
3	N S M

Pluralität: N, S, M

Ordnung des Paares (M,N) ändert sich, obwohl jeder Wähler bzgl. M und N die gleiche paarweise Präferenz hat in beiden Instanzen.

Satz von Arrow

Satz (Arrow, 1950)

Jede soziale Nutzenfunktion über einer Menge $|A| \geq 3$ Kandidaten, die Einstimmigkeit und IIA erfüllt, ist eine Diktatur.

Sei F eine soziale Nutzenfunktion mit den genannten Eigenschaften (Einstimmigkeit und IIA).

Lemma (Pairweise Neutralität)

Seien $\succ_1, \ldots, \succ_n$ und $\succ'_1, \ldots, \succ'_n$ zwei Präferenzprofile, und $\succ = F(\succ_1, \ldots, \succ_n)$ und $\succ' = F(\succ'_1, \ldots, \succ'_n)$. Wenn für jeden Wähler i gilt $a \succ_i b \Leftrightarrow c \succ'_i d$, dann $a \succ b \Leftrightarrow c \succ' d$.

Beweis:
Wir benennen zuerst die Kandidaten um, so dass $a \succ b$ und $c \neq b$ (aber evtl. $a = c$, und/oder $b = d$).

Pairweise Neutralität

- Jetzt passen wir $\succ_i$ und $\succ_i'$ an, damit sie identisch werden bzgl. a, b, c, d. Wir bewegen c und d in $\succ_i$, sowie a und b in $\succ_i'$:

  $$
  \begin{array}{llcl}
  \succ_1: & ..., a, ..., b, ... & \rightarrow & ..., c, a, ..., b, d, ... \\
  \succ_1': & c, ..., d, ... & \rightarrow & c, a, ..., b, d, ... \\
  \\
  \succ_2: & ..., b, ..., a, ... & \rightarrow & ..., b, d, ..., c, a, ... \\
  \succ_2': & ..., d, c, ... & \rightarrow & ..., b, d, c, a, ...
  \end{array}
  $$

  und so weiter...
  (Beachte: Per Annahme gilt $a \succ_i b \Leftrightarrow c \succ_i' d$)

- IIA garantiert, dass a und b in der gleichen Ordnung bleiben in $\succ$; c und d bleiben in der gleichen Ordnung in $\succ'$. Ebenfalls mit IIA folgt, dass wir nun alle anderen Elemente beliebig umstellen können. Dann gilt $\succ_i' = \succ_i$.
- Mit Einstimmigkeit gilt nun $c \succ a$ und $b \succ d$, so $c \succ d$. Mit $\succ_i = \succ_i'$ für alle i erhalten wir auch $c \succ' d$. $\square$ (Lemma)

Wer ist der Diktator?

Beweis (Satz):
Pairweise Neutralität zeigt, dass eine soziale Nutzenfunktion, die Einstimmigkeit und IIA erfüllt, einen allgemein zugrunde liegenden Ansatz verfolgt, die globale Präferenz zu bestimmen. Dieser Ansatz ist ähnlich für alle genannten Ordnungen und alle pairweisen Vergleiche zwischen Kandidaten. Diese Einsicht kann man ausnutzen, um zu zeigen, dass die Funktion sogar eine Diktatur ist.

Sei $a \neq b$ und $c \neq d$.

- Wenn es keine Wähler mit $a \succ_i b$ gibt, dann ist $b \succ a$.
- Wenn es nur Wähler mit $a \succ_i b$ gibt, dann ist $a \succ b$.
- Kritischer Punkt: i^* Wähler

1	...	$i^* - 1$	i^*	...	n	Resultat
$a \succ_i b$			$b \succ_i a$			$b \succ a$
$a \succ_i b$				$b \succ_i a$		$a \succ b$

Behauptung: i^* ist der Diktator!

i^* ist der Diktator

- i^* ist ein Diktator wenn $c \succ_{i^*} d \Rightarrow c \succ d$ für alle $c \neq d \in A$.
- Betrachte eine beliebige Menge an Präferenzen mit $c \succ_{i^*} d$ und $e \in A$ mit $e \neq c$ und $e \neq d$.
- Verschiebe drittes Element e, so dass es geordnet wird in $\succ_i$ wie folgt:

<table>
<tr><td>1</td><td>e</td><td colspan="6">...</td></tr>
<tr><td>...</td><td>e</td><td colspan="6">...</td></tr>
<tr><td>i^*</td><td>...</td><td>c</td><td>...</td><td>e</td><td>...</td><td>d</td><td>...</td></tr>
<tr><td>...</td><td colspan="6">...</td><td>e</td></tr>
<tr><td>n</td><td colspan="6">...</td><td>e</td></tr>
</table>

- Wegen IIA ändert das nicht die Ordnung von c und d in $\succ$.
- (c, e) erscheint genau wie vorher (a, b). Mit paarweiser Neutralität gilt $c \succ e$. Genauso erhalten wir $e \succ d$.
- Damit $c \succ d$. Da c und d beliebige Elemente waren, bestimmt die Präferenz von i^* also das ganze Ergebnis. $\square$ (Satz)

Eigenschaften von sozialen Auswahlfunktionen

- f kann strategisch manipuliert werden durch Wähler i wenn es $\succ_1, \ldots, \succ_n$ und $\succ'_i$ gibt mit $a \succ_i b$, wobei $b = f(\succ_1, \ldots, \succ_n)$ und $a = f(\succ_1, \ldots, \succ'_i, \ldots, \succ_n)$. f wird anreizkompatibel (incentive compatible, IC) genannt wenn es nicht strategisch manipuliert werden kann.
- f ist monoton wenn aus $f(\succ_1, \ldots, \succ_n) = a \neq b = f(\succ_1, \ldots, \succ'_i, \ldots, \succ_n)$ folgt, dass $a \succ_i b$ und $b \succ'_i a$.
- Wähler i ist ein Diktator in f: Für jede $\succ_1, \ldots, \succ_n \in L$ gilt, dass wenn $a \succ_i b$ für alle $b \neq a$, dann $f(\succ_1, \ldots, \succ_n) = a$. Dann ist f eine Diktatur.
- f ist surjektiv oder ausschöpfend in A wenn es für jeden Kandidaten $a \in A$ eine Menge von Präferenzen gibt, bei der a der Sieger ist.

Satz von Gibbard-Satterthwaite

Proposition

Eine soziale Auswahlfunktion f ist anreizkompatibel genau dann wenn f monoton ist.

Beweis: Direkte Folgerung aus der Definition. □

Satz (Gibbard 1973; Satterthwaite 1975)

Sei f eine soziale Auswahlfunktion f, die surjektiv ist mit $|A| \geq 3$. f ist anreizkompatibel genau dann wenn f eine Diktatur ist.

Beweis:
Wir beweisen die nicht-triviale Richtung des Satzes. Für jede monotone soziale Auswahlfunktion f, die ausschöpfend in A ist, leiten wir eine soziale Nutzenfunktion F her, die Einstimmigkeit und IIA erfüllt.

Erweiterung von sozialen Auswahlfunktionen

Sei $\succ$ eine Präferenzordnung und $S \subset A$. Wir nennen $\succ^S$ die Ordnung, in der alle Elemente von S in ihrer Ordnung in $\succ$ nach vorn geschoben werden.

Beispiel: $S = \{a, b, c\}, \quad A = S \cup \{d, e, f\}$

			$\succ$			$\rightarrow$			$\succ^S$			
a	e	d	c	f	b	$\rightarrow$	a	c	b	e	d	f
b	f	e	d	a	c	$\rightarrow$	b	a	c	f	e	d

Sei F eine f-erweiternde soziale Nutzenfunktion mit $F(\succ_1, \ldots, \succ_n) = \succ$, wobei $a \succ b$ genau dann wenn $f(\succ_1^{\{a,b\}}, \ldots, \succ_n^{\{a,b\}}) = a$.

Es ist erstmal gar nicht klar, ob F auf diese Weise überhaupt eine konsistente Ordnung $\succ$ erzeugt, also eine gültige soziale Nutzenfunktion darstellt.

Beweis durch Widerspruch

Lemma

Wenn f eine surjektive, anreizkompatible soziale Auswahlfunktion ist, dann ist die Erweiterung F eine soziale Nutzenfunktion.

Zeige Antisymmetrie und Transitivität.

Lemma

Wenn f eine surjektive anreizkompatible soziale Auswahlfunktion und keine Diktatur ist, dann erfüllt die Erweiterung F Einstimmigkeit, IIA und ist auch keine Diktatur.

Das ergibt einen Widerspruch mit dem Satz von Arrow.

Beweis der Lemmas

Wir beweisen den Satz durch Nachprüfen der Eigenschaften von F:

- Antisymmetrie: Wenn $a \succ b$ und $b \succ a$, dann $a = b$.
- Transitivität: Wenn $a \succ b$ und $b \succ c$, dann $a \succ c$.

- Einstimmigkeit: $F(\succ, \ldots, \succ) = \succ$.
- IIA
- Keine Diktatur

Eigenschaften

Behauptung

Für jedes $\succ_1, \ldots, \succ_n$ und jedes S ist der Sieger $f(\succ_1^S, \ldots, \succ_n^S) \in S$.

Beweis:

- f ist ausschöpfend, also gibt es $\succ_1'', \ldots, \succ_n''$ mit einem Sieger $a \in S$.
- Bewege nacheinander die Elemente von S nach vorne, sortiere die Elemente in $A \setminus S$ hinten um, sortiere die Elemente in S vorne um

$\Rightarrow$ Eine Transformation in $\succ_1^S, \ldots, \succ_n^S$.

- Monotonie stellt sicher, dass kein $b \notin S$ jemals zum Sieger wird während dieser Transformation. $\square$ (Behauptung)

Eigenschaften

- Antisymmetrie: Wenn $a \succ b$ und $b \succ a$, dann $a = b$.

  Gilt, da $f(\succ_1^{\{a,b\}}, ..., \succ_n^{\{a,b\}}) \in \{a, b\}$.

- Transitivität: Wenn $a \succ b$ und $b \succ c$, dann $a \succ c$.

  Für einen Widerspruch nehmen wir an, dass $a \succ b \succ c \succ a$.
  Wähle $S = \{a, b, c\}$. OBdA. sei $f(\succ_1^S, \ldots, \succ_n^S) = a$. Transformation zu $\succ^{S'}$ mit $S' = \{a, c\}$ zeigt $f(\succ_1^{S'}, \ldots, \succ_n^{S'}) = a$, und daher $a \succ c$. Dies ergibt einen Widerspruch mit Antisymmetrie.

Damit gezeigt: Wenn f anreizkompatibel und ausschöpfend in A ist, dann ist F eine gültige soziale Nutzenfunktion.

Eigenschaften

- Einstimmigkeit: $F(\succ, \ldots, \succ) = \succ$.

  Wenn $a \succ_i b$ für alle i, dann gilt mit Behauptung und Monotonie $f(\succ_1^{\{a,b\}}, \ldots, \succ_n^{\{a,b\}}) = a$.

- IIA:

  Wir nehmen an $a \succ_i b \Leftrightarrow a \succ'_i b$. Beachte $f(\succ_1^{\{a,b\}}, \ldots, \succ_n^{\{a,b\}}) = f(\succ_1'^{\{a,b\}}, \ldots, \succ_n'^{\{a,b\}})$, da bei der Transformation von $\succ_i^{\{a,b\}}$ in $\succ_i'^{\{a,b\}}$ das Ergebnis gleich bleibt aufgrund von Monotonie und der Behauptung.

- Keine Diktatur: Offensichtlich. $\square$ (Satz)

Präferenzen mit einem Scheitelpunkt

Der Satz von Gibbard-Satterthwaite ist entmutigend, aber er setzt voraus, dass die Wähler allgemeine Präferenzen haben können. Wenn die Präferenzen mehr Struktur beinhalten, dann existiert eine reichere Klasse von anreizkompatiblen Auswahlfunktionen.

Wir betrachten eine Menge von Alternativen, die entlang einer Linie geordnet werden können. Wir nehmen an, dass $A \subseteq [0, 1]$.

Definition

Eine Präferenzordnung $\succ_i$ über A hat einen Scheitelpunkt wenn es einen Punkt $p_i \in A$ gibt, so dass für alle $x \in A \backslash \{p_i\}$ und $\lambda \in [0, 1)$ gilt

$$(\lambda x + (1 - \lambda)p_i) \succ_i x \quad .$$

Als Beispiel betrachten wir das Problem, die Raumtemparatur in einem gemeinsamen Büro einzustellen. Jeder Angestellte hat eine optimale Temparatur und möchte, dass die Temparatur so nah wie möglich daran eingestellt wird.

Ordnungsmechanismen

Für Präferenzen mit einem Scheitelpunkt kann man den Satz von Gibbard-Satterthwaite nicht anwenden.

Ordnungsmechanismus für Präferenzen mit einem Scheitelpunkt:

- Sei k eine Zahl in $\{1, \ldots, n\}$
- Frage nur nach Scheitelpunkten $p_1, \ldots, p_n$ der Wähler.
- Sortiere die Scheitelpunkte von 1 nach 0 und gib den k-ten Scheitelpunkt in der Sortierung aus

Proposition

Für jedes feste $k \in \{1, \ldots, n\}$ ist der Ordnungsmechanismus anreizkompatibel. Wenn $n \geq 2$, dann ist er keine Diktatur.

Ordnungsmechanismen

Beweis:
Sei p das Ergebnis wenn alle Wähler die wahren Scheitelpunkt nennen. Wenn $p_i > p$, dann kann Wähler i das Ergebnis nicht durch $p_i' > p_i$ ändern. Wenn er einen Scheitelpunkt $p_i' \leq p$ lügt, dann kann nur ein schlechteres Ergebnis $p' \leq p$ entstehen. Das gleiche Argument kann man für $p_i < p$ anwenden. Keine Diktatur ist offensichtlich. $\square$

Der bekannteste Fall ist der Median-Mechanismus mit $k = \lfloor (n+1)/2 \rfloor$. Mit dem Durchschnitt der Scheitelpunkte $\sum_{i=1}^{n} p_i/n$ wäre der Mechanismus dagegen nicht anreizkompatibel.

Mit dem gleichen Argument wie oben bleibt der Ordnungsmechanismus anreizkompatibel wenn – zusätzlich zu den genannten Scheitelpunkten – eine beliebige Anzahl von vorher bekannten und festgelegten Ergebnissen $y_j \in [0, 1]$ hinzugenommen und mit sortiert werden. Der Mechanismus wählt dann den k-t-größten Scheitelpunkt von $\{p_1, \ldots, p_n, y_1, \ldots, y_m\}$.

Ordnungsmechanismen

Für jedes feste k ist der Ordnungsmechanismus anonym, d.h. er erfüllt

$$f(\succ_1, \ldots, \succ_n) = f(\succ'_1, \ldots, \succ'_n),$$

wenn $(\succ_1, \ldots, \succ_n)$ eine Permutation von $(\succ'_1, \ldots, \succ'_n)$ ist.

Satz (Moulin 1980; Ching 1997)

Seien p_i die genannten Scheitelpunkte. Eine soziale Auswahlfunktion f ist anreizkompatibel, surjektiv und anonym für Präferenzen mit einem Scheitelpunkt genau dann wenn f Ordnungsmechanismus über einer Menge $\{p_1, \ldots, p_n, y_1, \ldots, y_m\}$ ist, wobei $y_j \in [0, 1]$ feste Ergebnisse sind.

Das Resultat ist eine vollständige Charakterisierung für anonyme anreizkompatible soziale Auswahlfunktionen. Anonymität ist notwendig: Eine Diktatur ist kein Ordnungsmechanismus, aber surjektiv und anreizkompatibel (aber nicht anonym).

Hausallokation

Matching mit Präferenzen über Objekte

- n Spieler und n Häuser
- Annahme: Spieler i besitzt Haus i
  (nicht unbedingt notwendig, vereinfacht Analyse)
- Spieler i hat eine Präferenzordnung $\succ_i$ über Häuser
- Weise jedem Spieler eines der Häuser zu

Die Menge A der Ergebnisse sind alle vollständigen Paarungen von Häusern und Spielern. Ein Spieler hat aber nur eine Präferenz für das Haus, das er erhält. Damit sind alle Paarungen, in denen Spieler i das gleiche Haus bekommt, für ihn gleichwertig.

Anreizkompatible Mechanismen mit "guten" Eigenschaften?

Top-Trading-Cycles

Sei $G = (V, E)$ der gerichtete Graph:

- V Menge verbleibender Spieler mit ihren Häusern
- E Menge von gerichteten Kanten:

$$(i, \ell) \in E \Leftrightarrow \ell \in V \text{ besitzt bestes verbleibendes Haus für } i \in V$$

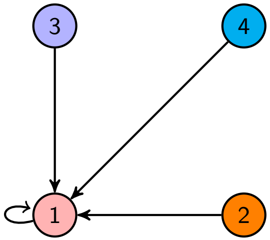
Top-Trading-Cycles (TTC) Mechanismus:

1. Frage Präferenzen der Spieler ab
2. **while** $V \neq \emptyset$
3. Erstelle Kantenmenge E wie oben beschrieben
4. Berechne alle gerichteten Kreise $C_1, \ldots, C_h$ in G (Schleifen gelten als Kreise, Kreise disjunkt)
5. **for** jede Kante (i, ℓ) in jedem Kreis $C_1, \ldots, C_h$ **do**
6. Weise Haus ℓ an Spieler i zu.
7. Entferne alle Spieler in $C_1, \ldots, C_h$ aus V

Beispiel

Spieler	Präferenzordnung			
1	1	2	3	4
2	1	3	2	4
3	1	4	2	3
4	1	2	4	3

Graph G in Runde 1:

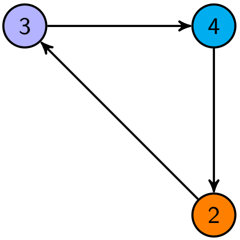


Zuweisung:
Spieler 1 – Haus 1

Beispiel

Spieler	Präferenzordnung		
2	3	2	4
3	4	2	3
4	2	4	3

Graph G in Runde 2:



Zuweisung:
Spieler 2 – Haus 3
Spieler 3 – Haus 4
Spieler 4 – Haus 2

Top-Trading-Cycles – Analyse

Beobachtungen:

- Jeder Spieler in G hat Ausgangsgrad 1. Es gibt immer mindestens einen Kreis in G, und alle Kreise in G sind disjunkt.
- Seien V_k die Spieler, die in der k-ten Runde von TTC entfernt werden. Jeder Spieler in V_k bekommt das Haus, das er am liebsten mag – abgesehen von Häusern der Spieler in $V_1 \cup \ldots \cup V_{k-1}$.
- Spieler $i \in V_k$ bekommt sein bestes Haus in der k-ten Runde. Der Besitzer dieses Hauses liegt auch in V_k und erhält somit auch sein bestes Haus.

Satz (Roth 1982)

Der TTC Mechanismus ist anreizkompatibel.

Beweis:
Betrachte Spieler $i \in V_j$ mit wahrer Präferenz $\succ_i$. Wenn er die Wahrheit sagt, bekommt er das beste verbleibende Haus in Runde j.

Top-Trading-Cycles – Analyse

Betrachte ein Haus von einem Spieler in $V_1 \cup \ldots \cup V_{j-1}$, das ihm besser gefällt. Spieler i kann keines dieser Häuser bekommen:

- In Runde $k = 1, \ldots, j-1$ will kein Spieler $\ell \in V_k$ das Haus von i haben, sonst wäre i mit ℓ in einem Kreis.
- Kein Spieler $\ell \in V_k$ möchte das Haus von i in einer Runde $< k$ haben, sonst würde ℓ es in Runde k immer noch haben wollen.

Die Häuser der Spieler in $V_1 \cup \ldots \cup V_{j-1}$ bleiben unerreichbar, egal welche Präferenzordnung Spieler i nennt. Wenn i aber bis Runde j warten muss, dann liefert TTC bei ehrlicher Bewertung das beste (verbleibende) Haus. $\square$

TTC ist also ein anreizkompatibler Mechanismus, aber davon gibt es mehrere.

Warum ist TTC besser als andere Ansätze?

Top-Trading-Cycles – Kernzuweisung

Sei M eine Zuweisung von Häusern – Spieler i bekommt Haus $M(i)$.
Sei M_S eine Zuweisung, die aus M entsteht, wenn die Koalition $S \subseteq \mathcal{N}$ ihre initialen Häuser anders unter sich aufteilen.

Definition

Eine Spielermenge $S \subseteq \mathcal{N}$ bildet eine blockierende Koalition für M wenn es eine Zuweisung M_S gibt, so dass

- für jeden Spieler $j \in C$ ist M_C genauso gut: $M_C(j) \succeq_j M(j)$
- für mind. einen Spieler $i \in C$ ist M_C besser: $M_C(i) \succ_i M(i)$

Eine Zuweisung M ohne blockierende Koalition liegt im Kern des Spiels.

Eine Kernzuweisung hat eine Optimalitätseigenschaft: Keine Teilmenge von Spielern möchte ihre Häuser vom Mechanismus zurückhalten und anders (unter sich) verteilen. Insbesondere bekommt jeder Spieler i ein Haus, das er mindestens so gerne mag wie sein anfängliches Haus i.

Top-Trading-Cycles – Kernzuweisung

Satz (Roth, Postlewaite 1977)

Der TTC Mechanismus berechnet die eindeutige Zuweisung im Kern des Spiels.

Beweis:
Induktion: Nur die TTC-Zuweisung kann im Kern sein, aber keine andere.

- Anfang: Jeder $i \in V_1$ bekommt sein allerbestes Haus. V_1 ist also blockierende Koalition für jede Zuweisung, die die Häuser in V_1 nicht so verteilt wie TTC.
- Annahme: Die Häuser der Spieler in $V_1, \ldots, V_{j-1}$ müssen verteilt werden wie bei TTC.
- Schritt: Unter der Annahme bekommt jeder $i \in V_j$ das beste (verbleibende) Haus. V_j ist also blockierende Koalition für jede Zuweisung, die die Häuser in $V_1, \ldots, V_{j-1}$ so verteilt wie TTC, aber V_j nicht so verteilt wie TTC.

Damit gilt: Entweder die TTC-Zuweisung ist im Kern oder der Kern ist leer.

Top-Trading-Cycles – Kernzuweisung

Sei nun M die TTC-Zuweisung und M_S eine Umverteilung der Häuser der Spieler $S \subseteq M$. Spieler $i \in S$ hat vorher Haus i. Wenn S mit M_S blockierend sein soll, dann muss i nachher ein Haus erhalten, sonst wäre er schlechter dran als in M. Damit bildet die Umverteilung in S eine Menge von Kreisen:

- Kreis enthält Spieler von N_j und N_ℓ mit $j < \ell$:
  Mindestens ein Spieler $i \in N_j$ bekommt ein Haus von einem Spieler in N_ℓ und ist damit schlechter dran als in M.
- Kreis enthält nur Spieler einer Menge N_j:
  In M bekommt jeder Spieler $i \in N_j$ sein bestes Haus unter den Häusern der Spieler in N_j. Mindestens ein Spieler ist schlechter dran als in M.

Also ist die TTC-Zuweisung im Kern des Spiels, und somit die eindeutige Zuweisung mit dieser Eigenschaft. $\square$

Random Serial Dictatorship

Random Serial Dictatorship (RSD) Mechanismus:

1. Ordne Spieler in zufälliger Reihenfolge
2. Frage Präferenzen der Spieler ab
3. Sei V die Menge aller Häuser
4. **for** $i = 1, 2, 3, \ldots, n$ **do**
5. Weise Spieler i sein bestes Haus h aus V zu
6. Entferne h aus V

Dieser Mechanismus ist eine geordnete Variante der Diktatur.
Das folgende Resultat kann man ähnlich wie für TTC beweisen.

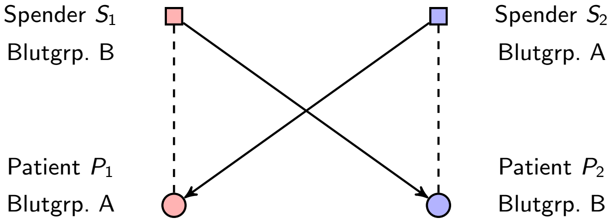
Satz

Für jede gewählte Ordnung ist der RSD Mechanismus anreizkompatibel.

Liegt die Zuweisung von RSD im Kern? Für jede Ordnung? Für keine? Für einige?

Nierenspende

In vielen Ländern gibt es mittlerweile (Pläne für) Programme zum Nierentausch. Patienten haben oft einen Bekannten/Verwandten als Lebendspender, dessen Niere (z.B. aufgrund von Blutgruppe) nicht zum Patienten passt. Das Ziel ist ein Organaustausch: Zwei inkompatible (Patient,Spender)-Paare tauschen die Spenderorgane aus, wenn sie zum jeweils anderen Patienten passen.



Top-Trading-Cycles für Nierenspende

Im Prinzip also eine Hausallokation:

- "Häuser" sind Spenderorgane, "Spieler" sind Patienten
- Präferenzen über Organe = W.keit der erfolgreichen Transplantation.
- TTC Mechanismus ist anreizkompatibel und im Kern des Spiels

Anmerkungen:

- Keine legale Verpflichtung zur Spende. Alle Operationen eines Kreises werden simultan durchgeführt, da sonst ein Anreizproblem entsteht: Spender S_i tritt zurück sobald Patient P_i seine Niere erhalten hat.
- Schwierig: lange Kreise = viele simultane Operationen. Wenn dagegen eine reine Nierenspende dazukommt, wird der Kreis zum Pfad – dann wird immer erst die Niere von Spender S_i entnommen bevor Patient P_i seine Spende bekommt.
- Sehr oft eher binäre Präferenzen: Niere passt zum Patienten oder nicht.

Organspende durch Matching

Ansatz mit Matching in einem Graphen $G = (V, E)$:

- Für Patient P_i sei E_i die Menge von kompatiblen Spendern
- Jedes Paar (P_i, S_i) ist ein Knoten $v_i \in V$
- Wir suchen einfache Austausche bzw. Kreise der Länge 2:
  Kante $\{v_i, v_j\} \in E$ wenn $S_i \in E_j$ und $S_j \in E_i$
- Patient kann bei Nennung der Menge E_i lügen
- Offensichtlich: Es lohnt sich nur ein $F_i \subseteq E_i$ zu lügen

Matching-Mechanismus für Nierentausch:

- Frage die Mengen F_i der Patienten ab
- Erstelle Graph G wie oben, wobei $E = \{\{v_i, v_j\} \mid S_i \in F_j, S_j \in F_i\}$
- Berechne ein Matching M von G mit größter Kardinalität
- Nierentausch gemäß der Kanten in M

Maximum Matching mit Prioritätsliste

Das Maximum Matching ist nicht eindeutig. Verschiedene maximum Matchings verteilen die Nieren an unterschiedliche Patienten. Wir müssen das Matching auf eine "monotone" Art wählen. Dafür priorisieren wir die Patienten. Das ist ein üblicher Ansatz, z.B. bei der Erstellung von Wartelisten für die Organspende.

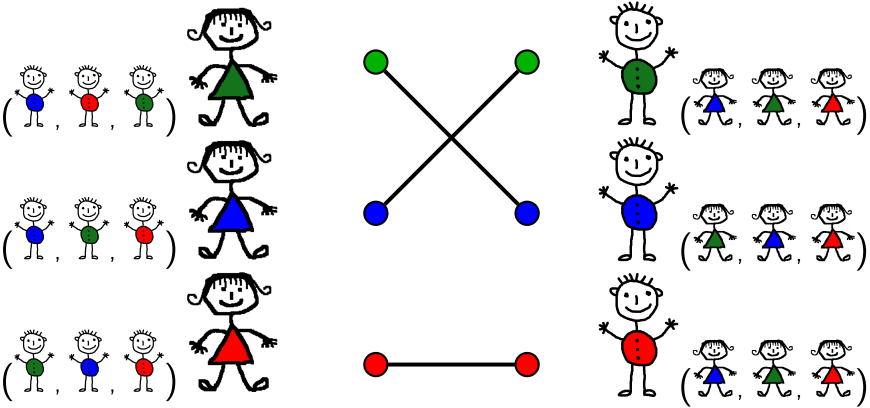
Maximum Matching mit Prioritäten

1. Sei M_0 die Menge aller maximum Matchings von G
2. **for** $i = 1, 2, 3, \ldots, n$ **do**
3. Sei Z_i die Menge von Matchings in M_{i-1}, in denen Knoten v_i gematcht ist
4. **if** $Z_i \neq \emptyset$ **then** $M_i \leftarrow Z_i$; **else** $M_i \leftarrow M_{i-1}$.
5. **return** beliebiges Matching aus M_n

Satz

Der Matching-Mechanismus mit Prioritäten für Nierenaustausch ist anreizkompatibel.

Stabiles Matching



Jeder Spieler hat eine Präferenzordnung über die Spieler der anderen Seite
Ordnung: (bester Partner, zweitbester, ...).

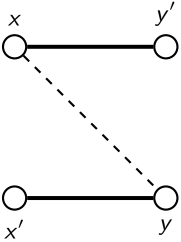
Stabiles Matching

- Menge $\mathcal{X}$ von n Männern, Menge $\mathcal{Y}$ von n Frauen
- Jeder Mann $x \in \mathcal{X}$ hat Präferenzordnung $\succ_x$ über alle $y \in \mathcal{Y}$.
- Jede Frau $y \in \mathcal{Y}$ hat Präferenzordnung $\succ_y$ über alle $x \in \mathcal{X}$.
- Jede Person ist lieber verparntnert als allein.
- Für ein Matching M sei $M(x) \in \mathcal{Y}$ die Partnerin von Mann $x \in \mathcal{X}$ in M, und $M(y) \in \mathcal{X}$ der Partner von Frau $y \in \mathcal{Y}$ in M.
- Sei $M(x) = *$ wenn x allein in M, und $M(y) = *$ genauso.

Stable Matching

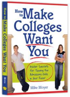
Wann ist ein Matching stabil? Was ist eine Gefahr für Stabilität?

- In M ist $\{x, y\}$ ein blockierendes Paar genau dann wenn x und y sich gegenseitig lieber mögen als ihre jeweiligen Partner $y' = M(x)$ und $x' = M(y)$.
- Ein Matching M ohne blockierendes Paar ist ein stabiles Matching.



Stabiles Matching spielt eine entscheidende Rolle in vielen Anwendungen:

Ärzte/Krankenhäuser Uni-Zulassung Arbeitsmarkt etc.



Existenz und Berechnung

Deferred-Acceptance Algorithmus (mit Männerantrag):

1. **while** es gibt ungematchten Mann $x \in \mathcal{X}$ **do**
2. $\quad$ x macht der liebsten Frau y einen Antrag, die ihn noch nicht ablehnt
3. $\quad$ **if** y ungematcht **then** $\{x, y\}$ vorläufig gematcht;
4. $\quad$ **else if** y vorläufig gematcht mit x' **then**
5. $\quad\quad$ **if** $x \succ_y x'$ **then** $\{x, y\}$ vorläufig gematcht, y lehnt x' ab;
6. $\quad\quad$ **else** $\{x', y\}$ vorläufig gematcht, y lehnt x ab;

Satz (Gale, Shapley 1962)

Der Deferred-Acceptance Algorithmus berechnet ein stabiles Matching nach höchstens n^2 Iterationen.

Beweis:
Der Algorithmus braucht nur n^2 Iterationen, da jeder Mann jeder Frau höchstens einmal einen Antrag macht. Jeder Spieler ist zu jeder Zeit mit maximal einem Partner vorläufig gematcht. Daher berechnet der Algorithmus ein Matching M.

Existenz und Berechnung

Im Ablauf des Algorithmus gelten die folgenden Invarianten:

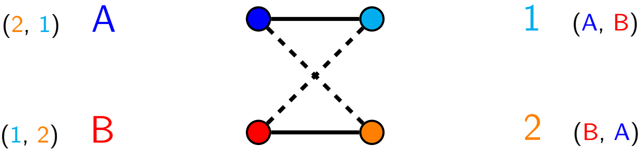
1. Für jeden Mann x ist die Präferenz der Frauen, denen er einen Antrag macht, streng monoton fallend.
2. Für jede Frau y ist die Präferenz des vorläufigen Partners streng monoton steigend. Damit gilt insbesondere, wenn eine Frau vorläufig gematcht ist, dann ist sie auch im endgültigen Matching gematcht.

Nehmen wir nun an, M hat ein blockierendes Paar $\{x, y\}$ with $y \succ_x M(x)$ und $x \succ_y M(y)$. x mag seine Partnerin $M(x)$ weniger (oder ist womöglich sogar alleine). Wegen Invariante 1. hat er y einen Antrag gemacht und wurde abgelehnt. Also muss y einen Antrag von einem besseren Mann x' erhalten haben. Wegen Invariante 2. kann ihr endgültiger Partner in M nur (noch) besser sein. Also, $M(y) \succeq_y x' \succ_y x$, ein Widerspruch. $\square$

Eindeutigkeit

Der Deferred-Acceptance Algorithmus ist unterspezifiziert: Es wird nicht gesagt, welcher Mann als nächstes einen Antrag macht.

- Gibt es mehrere stabile Matchings?
- Kann der Algorithmus mehrere (alle?) stabile Matchings berechnen?



Stabile Matchings sind nicht eindeutig. In diesem Fall gibt es ein männeroptimales Matching (durchgezogen) und ein frauenoptimales (gestrichelt), bei dem jede(r) gleichzeitig den(die) beste(n) Partner(in) bekommt.

Optimalität

Sei $h(x) \in \mathcal{Y}$ die beste Frau, so dass $\exists$ stabiles Matching M' mit $M'(x) = h(x)$.
Sei $h(y) \in \mathcal{X}$ der beste Mann, so dass $\exists$ stabiles Matching M' mit $M'(y) = h(y)$.

Definition

Ein stabiles Matching M ist

- männeroptimal wenn $M(x) = h(x)$ für alle $x \in \mathcal{X}$.
- frauenoptimal wenn $M(y) = h(y)$ für alle $y \in \mathcal{Y}$.

In einem optimalen Matching bekommen alle Spieler einer Seite gleichzeitig den besten Partner, die sie in irgendeinem stabilen Matching bekommen können. Es ist nicht klar, ob so etwas überhaupt möglich ist. Der folgende Satz beweist dies, und sogar noch mehr: Egal in welcher Reihenfolge die Anträge im Algorithmus gemacht werden, es kommt immer ein eindeutiges optimales Matching heraus.

Satz

Der Deferred-Acceptance Algorithmus mit Männerantrag berechnet immer genau das männeroptimale stabile Matching und mit Frauenantrag das frauenoptimale.

Optimalität

Beweis: Mit Widerspruch zur Stabilität:

- Sei M das Matching des Algorithmus mit Männerantrag.
- Sei M' stabiles Matching, in dem ein Mann $j \in \mathcal{X}$ eine strikt bessere Partnerin $M'(j) \succ_j M(j)$ bekommt.
- Da $M'(j) \succ_j M(j)$, gibt es eine Iteration im Algorithmus, in der j der Frau $M'(j)$ einen Antrag gemacht hat und abgelehnt wird.
- Eventuell gibt es noch mehr Iterationen im Algorithmus, in denen ein Mann x von der Partnerin $M'(x)$ in M' abgelehnt wird.
- Betrachte die erste dieser Iterationen. Frau $M'(x)$ lehnt Mann x nur ab, weil sie einen besseren Antrag von Mann $i \neq x$ bekommt (d.h. $i \succ_{M'(x)} x$).
- Da dies die allererste Ablehnung dieser Art ist, mag Mann i die Frau $M'(x)$ mehr als $M'(i)$ (d.h. $M'(x) \succ_i M'(i)$).
- Also ist $(i, M'(x))$ ein blockierendes Paar in M' und M' kein stabiles Matching – Widerspruch $\square$

Anträge und Anreize

Satz

Der Deferred-Acceptance Algorithmus mit Männerantrag ist anreizkompatibel für die Männer.

Beweis:
Wir vereinbaren folgende Notation. Ein Lügner habe oBdA die Nummer 1.

- Echtes Profil: $\pi = (\succ_1, \ldots, \succ_n)$, Algorithmus berechnet M
- Mann 1 lügt: $\pi' = (\succ', \succ_2, \ldots, \succ_n)$, Algorithmus berechnet M'

Wenn Lügen profitabel ist, dann $M'(1) \succ_1 M(1)$. Wir werden zeigen, dass in diesem Fall M' nicht stabil ist für π' – ein Widerspruch.

Die folgende Behauptung zeigt, dass in der Menge der Männer, die von der Lüge profitieren, die zugewiesenen Partnerinnen untereinander ausgetauscht werden.

Anträge und Anreize

Behauptung

Seien $R = \{x \mid M'(x) \succ_x M(x)\}$ die Männer, die in M' profitieren. Für jeden Mann $x \in R$ und seine neue Partnerin $y = M'(x)$ ist der Partner $x' = M(y)$ in M auch $x' \in R$.

Beweis:

- $x' = 1$: Klar, $x' \in R$ per Annahme, dass 1 lügen möchte
- $x' \neq 1$: Da $x \in R$ gilt $y \succ_x M(x)$. Dann $x' \succ_y x$, denn sonst hätte M das blockierende Paar $\{x, y\}$.
- Also: $M'(x') \succ_{x'} y$, denn sonst hätte M' das blockierende Paar $\{x', y\}$.
- Also: $M'(x') \succ_{x'} M(x')$ und somit $x' \in R$. $\square$ (Behauptung)

Die Menge der Partnerinnen von profitierenden Männern ist also in beiden Matchings M und M' die gleiche: $T = \{y \mid M(y) \in R\} = \{y \mid M'(x) \in R\}$.

Anträge und Anreize

Da alle Männer aus R besser gematcht sind in M', sind alle Frauen aus T schlechter gematcht, sonst gäb es ein blockierendes Paar in M.

Betrachte nun die Berechnung von M mit dem echten Profil π:

- Sei x der letzte Mann von R, der einen Antrag macht.
- Dieser Antrag geht an seine endgültige Partnerin $y = M(x) \in T$.
- Keine weiteren Anträge von Männern in R, daher alle außer x schon gematcht wie in M. Also: y hat $M'(y)$ in einer Runde davor abgelehnt.
- y kann nur einen Antrag eines Mannes $x' \notin R$ haben mit $x' \succ_y M'(y)$.
- $x' \notin R$ bedeutet $M(x') \succeq_{x'} M'(x')$, und Ablehnung von y bedeutet $y \succ_{x'} M(x')$.
- Somit $y \succ_{x'} M'(x')$ und $x' \succ_y M'(y)$.
- Da $x' \notin R$, gilt $x' \neq 1$. Also hat M' ein blockierendes Paar für π', ein Widerspruch. $\square$ (Satz)

Nicht anreizkompatibel für die passive Seite

Deferred-Acceptance mit Frauenantrag ist nicht anreizkompatibel für Männer:

	$\boxed{B}$	~~A~~	$\boxed{A}$
	A	$\boxed{C}$	B
	C	B	C
Frauen:	1	2	3

~~B~~	~~A~~	~~A~~
$\boxed{A}$	$\boxed{C}$	$\boxed{B}$
C	B	C
1	2	3

Männer:	A	B	C
	1	3	1
	$\boxed{3}$	$\boxed{1}$	3
	~~2~~	2	$\boxed{2}$

A	B	C
$\boxed{1}$	$\boxed{3}$	1
~~2~~	~~1~~	3
~~3~~	2	$\boxed{2}$

Literatur

- Arrow. A Difficulty in the Concept of Social Welfare. 1950.
- Gibbard. Manipulation of Voting Schemes. A general result. Econometrica 41(4):587-601, 1973.
- Satterthwaite. Strategy-proofness and Arrows Conditions. Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. Journal of Economic Theory 10(2):187-217. 1975.
- Moulin. On strategy-proofness and single peakedness. Public Choice 35(4):437–455, 1980.
- Ching. Strategy-proofness and "median voters". International Journal of Game Theory 26(4):473-490, 1997.
- Roth. Incentive compatibility in a market with indivisible goods. Economic Letters 9(2):127–132, 1982.
- Shapley, Scarf. On Cores and Indivisibility. Journal of Mathematical Economics 1(1):23–27, 1974.

Literatur

- Roth, Postlewaite. Weak versus strong domination in a market with indivisible goods. Journal of Mathematical Economics 4(2):131–137, 1977.
- Roth, Sönmez, Ünver. Pairwise kidney exchange. Journal of Economic Theory 125(2):151–188, 2005.
- Gale, Shapley. College Admissions and the Stability of Marriage. American Mathematical Monthly 69:914, 1962.
- Dubins, Freedman. Machiavelli and the Gale-Shapley Algorithm. American Mathematical Monthly 88(7):485–494, 1981.
- Roth. The Economics of Matching: Stability and Incentives. Mathematics of Operations Research 7(4):617–628, 1982.

- Brandt, Conitzer, Endriss, Lang, Procaccia. Handbook of Computational Social Choice (Kapitel 14)
- Nisan, Vazirani, Roughgarden, Tardos. Algorithmic Game Theory, 2007. (Kapitel 9 und 10)
- Roughgarden. Twenty Lectures on Algorithmic Game Theory, 2016. (Kaptitel 9.4 und 10)