Decentralized learning in control and optimization for networks and dynamic games

Part IV: Dynamics in games

Alexandre Proutiere

KTH

This part

- Aims at understanding how players may adapt their actions in repeated games

- Aims at modeling ***natural*** and ***robust*** ways of adapting actions over time, and at understanding the resulting dynamics

Some relevant books

- ***Strategic learning and its limits***
  H.P. Young, Oxford Univ. Press, 2004
- ***The theory of learning in games***
  D. Fudenberg and D. Levine, MIT Press 2004
- ***Evolutionary games and Equilibrium selection***
  L. Sammuelson, MIT Press, 1997
- ***Evolutionary game theory***
  J. Weibull, MIT Press, 1995
- ***Prediction, Learning, and Games***
  N. Cesa-Bianchi and G. Lugosi, Cambridge Univ. Press, 2006
- ***Learning, regret minimization, and equilibria***
  A. Blum and Y. Mansour, Chapter 4 in "Algorithmic Game Theory", Cambridge Univ. Press, 2007

Outline of Part IV

- Games and equilibria
- Nash dynamics
- Fictitious play
- No-regret dynamics
- Trial and error learning

Games

- A set of m agents or players
- Finite strategy set for player i: S_i
- Cost function for player i: $c_i : S = (S_1, \ldots, S_m) \rightarrow \mathbb{R}$

- Notation: $s = (s_1, \ldots, s_m) = (s_i, s_{-i})$

Ex 1: coordination game

- Coordination game

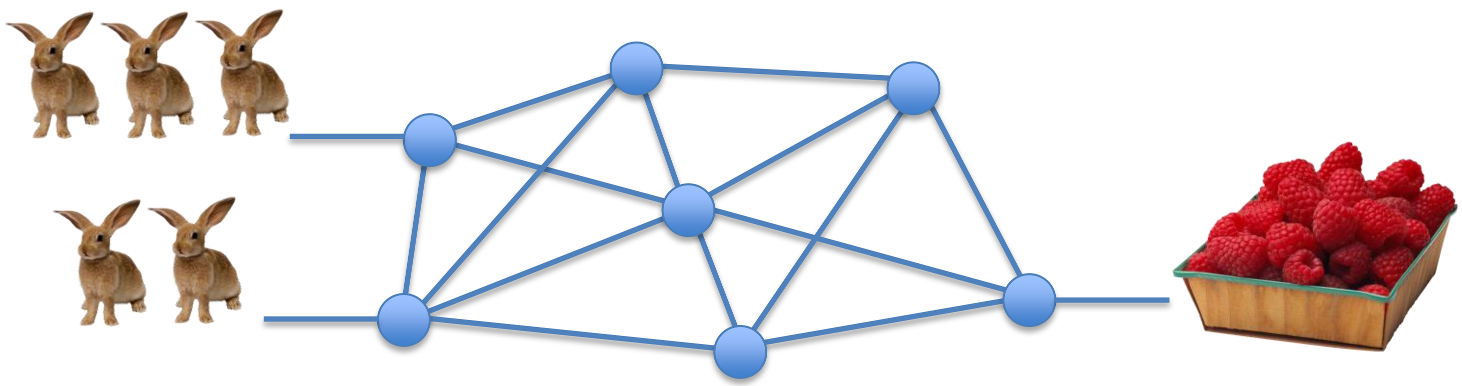
		Player 2 a	b
Player 1	A	(1,1)	(0,0)
	B	(0,0)	(1,1)

Ex 2: Shapley game

- Shapley game: pay-off matrix

		Player 2		
		L	M	R
	T	(0,0)	(1,0)	(0,1)
Player 1	M	(0,1)	(0,0)	(1,0)
	B	(1,0)	(0,1)	(0,0)

Network congestion game



- Network: set of links with limited capacity
- Strategies: set of routes to destination
- Latency function of link e: $l_e : \mathbb{N} \to \mathbb{R}_+$
- Under strategies $s : n_e(s)$ = number of users going through e
- Cost for user using route $r : \sum_{e \in r} l_e(n_e(s))$

Pure Nash Equilibrium

- A pure Nash equilibrium is a set of strategies $s = (s_1, \ldots, s_m)$ such that no player has incentive to modify her strategy

$$\forall i, \quad c_i(s'_i, s_{-i}) \geq c_i(s), \quad \forall s'_i \in S_i$$

- Nash equilibria are stable

Ex 1: coordination game

- Coordination game: pay-off matrix

		Player 2	
		a	b
Player 1	A	(1,1)	(0,0)
	B	(0,0)	(1,1)

Ex 2: Shapley game

- Shapley game
- No pure NE

Player 2

		L	M	R
Player 1	T	(0,0)	(1,0)	(0,1)
	M	(0,1)	(0,0)	(1,0)
	B	(1,0)	(0,1)	(0,0)

Mixed strategies

- A mixed strategy for player i is a distribution over S_i
- Set of mixed strategies: ΔS_i

$$p_i \in \Delta S_i, \quad p_i : S_i \to [0,1], \quad \sum_{s_i \in S_i} p_i(s_i) = 1$$

- Costs under $p = (p_1, \ldots, p_m) \in \Delta = \Delta_1 \times \ldots \Delta_m$

$$C_i(p) = \sum_{s=(s_1,\ldots,s_m)} p_1(s_1)...p_m(s_m)c_i(s)$$

Mixed Nash equilibrium

- $p = (p_1, \ldots, p_m) \in \Delta S = \Delta S_1 \times \ldots \Delta S_m$ is a mixed NE if:

$$\forall i, \quad C_i(p'_i, p_{-i}) \geq C_i(p), \quad \forall p'_i \in \Delta S_i$$

- Every game has at least one mixed NE (Brouwer's theorem)
- A pure NE is also a mixed NE

Nash dynamics

Best responses

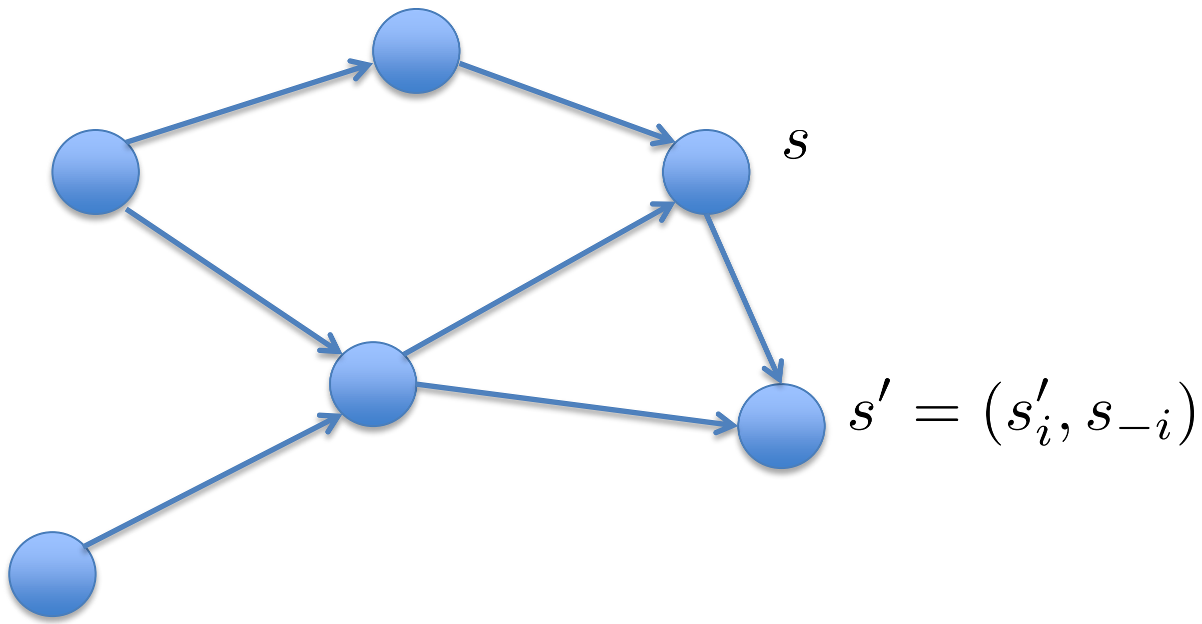
- Consider pure actions here
- Best response: a best response a_i against strategies s_{-i} is such that:

$$a_i \in \arg\min_{s_i \in S_i} c_i(s_i, s_{-i})$$

- Nash dynamics: a sequence of best responses (one player updates her strategy at a time)
- Liveness property: each player gets a chance of updating after at most a fixed number of updates
- Random Nash dynamics: players are chosen uniformly at random for updates

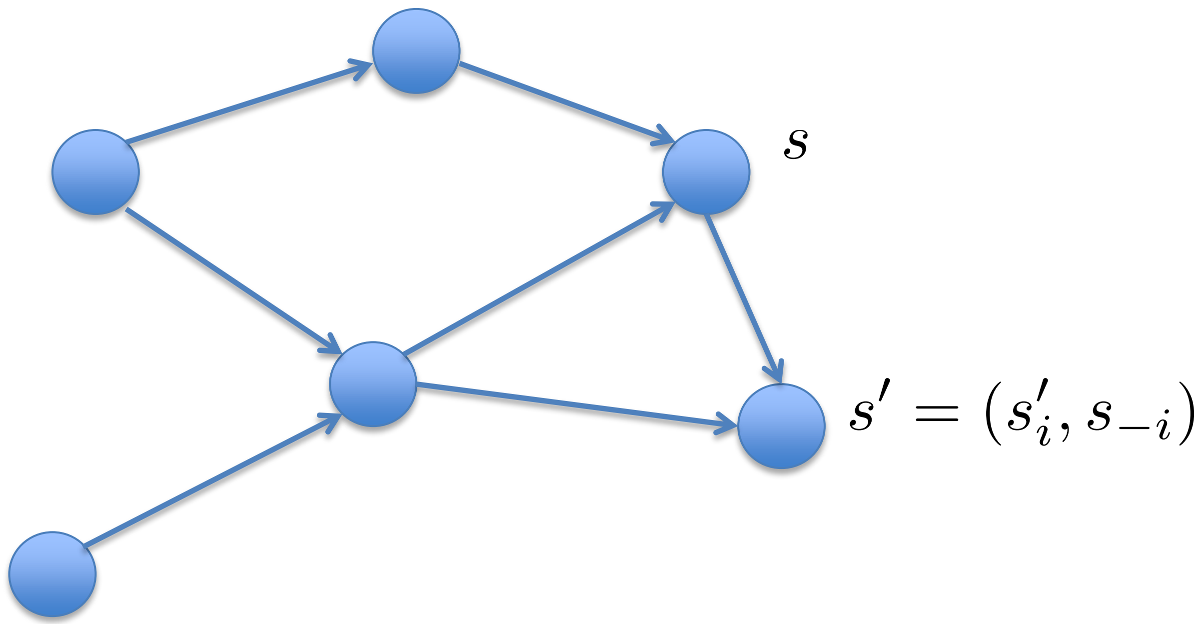
Graph representation

- Vertices: set of strategies
- Directed edges: best responses

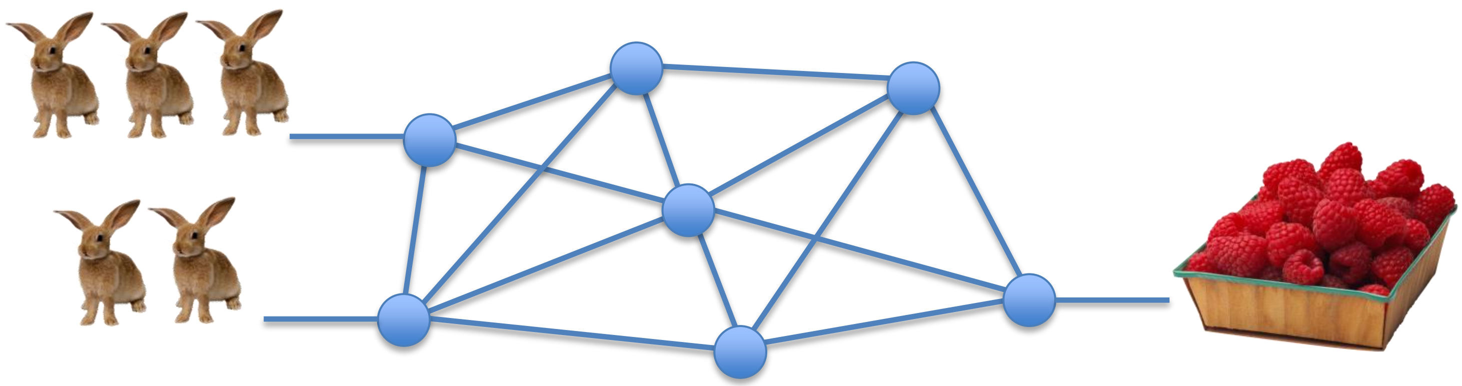


Graph representation

- Pure NEs = sinks of the graph



Network congestion game



- Network: set of links with limited capacity
- Strategies: set of routes to destination
- Latency function of link e: $l_e : \mathbb{N} \to \mathbb{R}_+$
- Under strategies $s : n_e(s)$ = number of users going through e
- Cost for user using route $r : \sum_{e \in r} l_e(n_e(s))$

Potential games

- Rosenthal, 1973
- Every network congestion game admits a potential function

$$s' = (s'_i, s_{-i})$$

$$\Phi(s) - \Phi(s') = c_i(s') - c_i(s)$$

$$0 \leq \Phi(s) \leq n.m.l_{\max}, \quad \forall s \in S_1 \times \ldots \times S_m$$

- Proof: $\Phi(s) = \sum_{e} \sum_{k=1}^{n_e(s)} l_e(k)$

- NEs are local minima of the potential function

Social efficiency of NEs

- There is a difference between NEs and socially optimal routing strategies:

NEs: $$\text{minimize} \quad \Phi(s) = \sum_{e} \sum_{k=1}^{n_e(s)} l_e(k)$$

Socially optimal routing:

$$\text{minimize} \quad \Phi(s) = \sum_{e} n_e(s) l_e(n_e(s))$$

Convergence of Nash dynamics

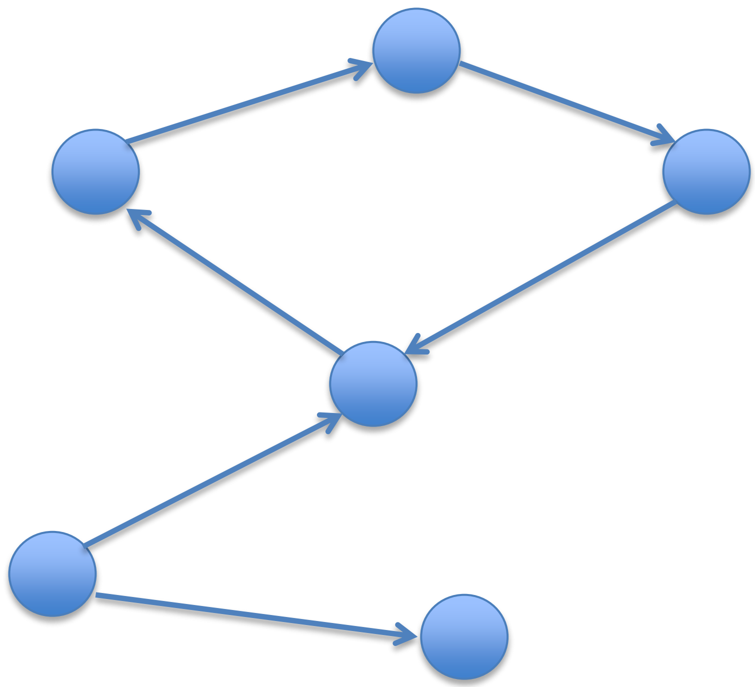
- Best response dynamics with liveness property converge to NEs
- Convergence time?

Theorem* There is a network congestion game and an intial condition such that all better response sequences have exponential (w.r.t. the number of players) length.

* The complexity of pure NEs, **Fabrikant-Papadimitriou-Talwar**, STOC, 2004

Non-potential games

- Notion of sink equilibrium*

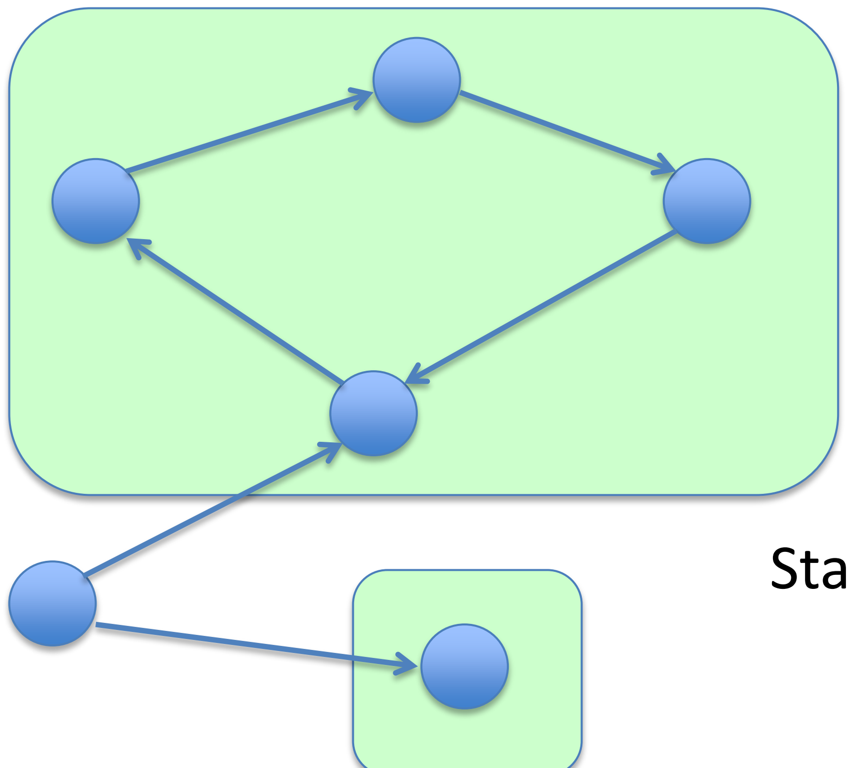


State graph

* **Goemans-Vetta**, FOCS, 2005

Non-potential games

- Notion of sink equilibrium*: strongly connected components without outgoing link

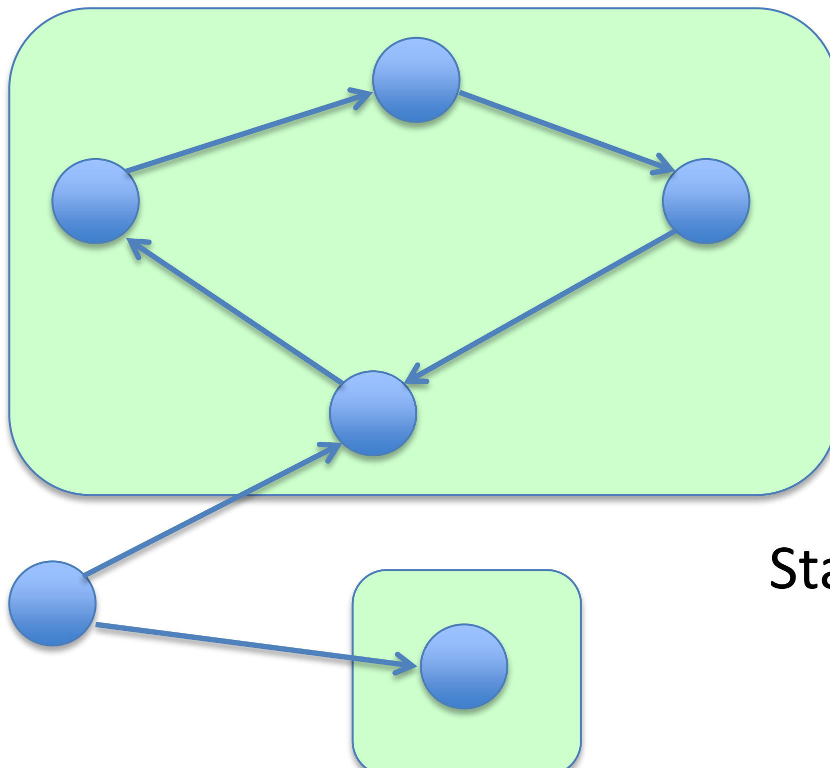


State graph

* **Goemans-Vetta**, FOCS, 2005

Non-potential games

- Every random Nash dynamics converge to a sink equilibrium
- Nothing else can be said



State graph

* **Goemans-Vetta**, FOCS, 2005

Stable marriage problem

- Two sets: set of women, set of men
- Each person has a preference list

A  (a,c,b)

B  (a,b,c)

C  (c,b,a)

a  (A,B,C)

b  (A,C,B)

c  (A,C,B)

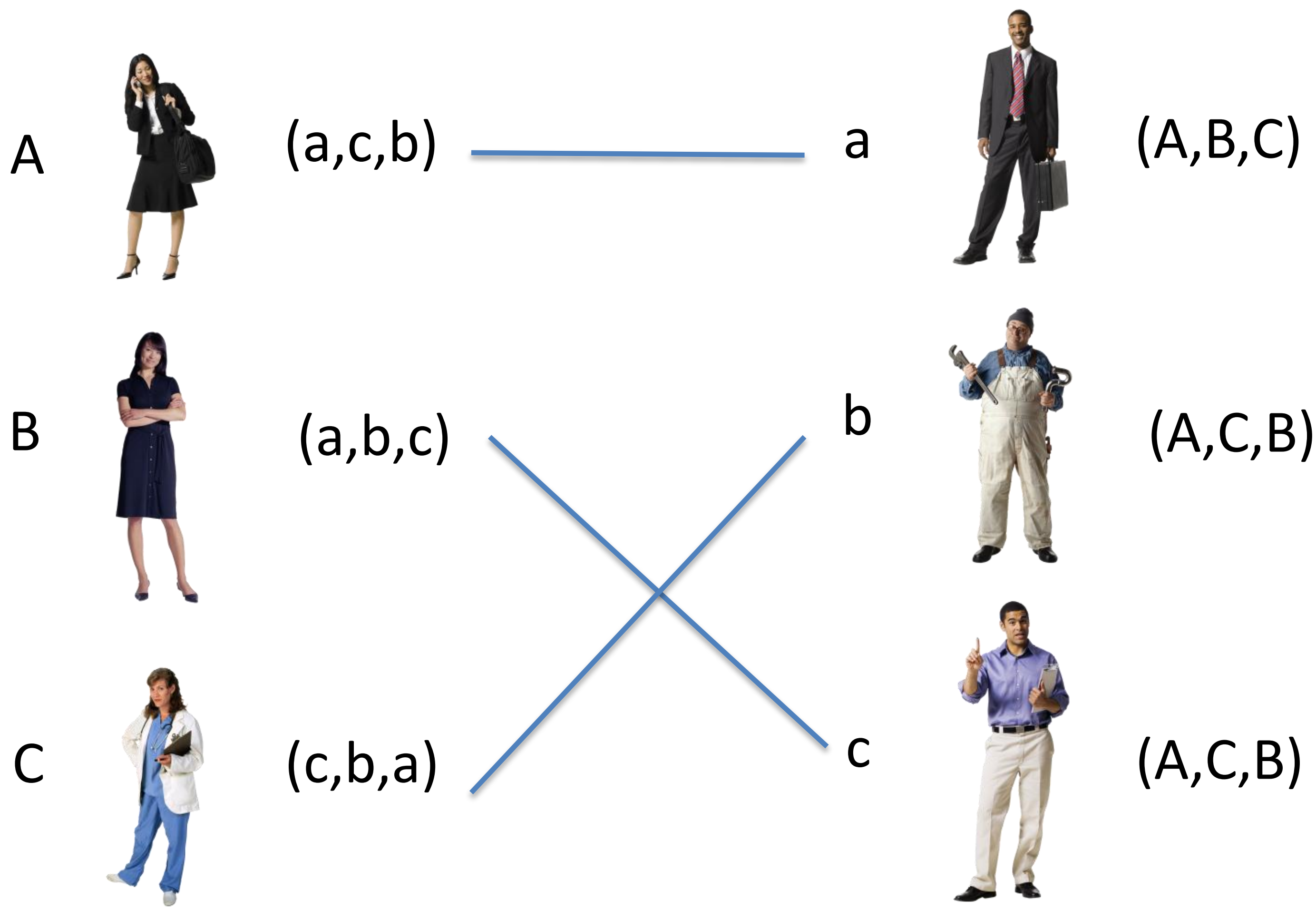
Applications

- Patients/hospitals
- Students/college
- Labor market
- …

- Connection to games: there is an active side (women) who proposes
  - Women are playing against each other
  - Strategy of a woman: proposes a single man, and gets the pay-off if she wins him
  - NEs = stable matchings

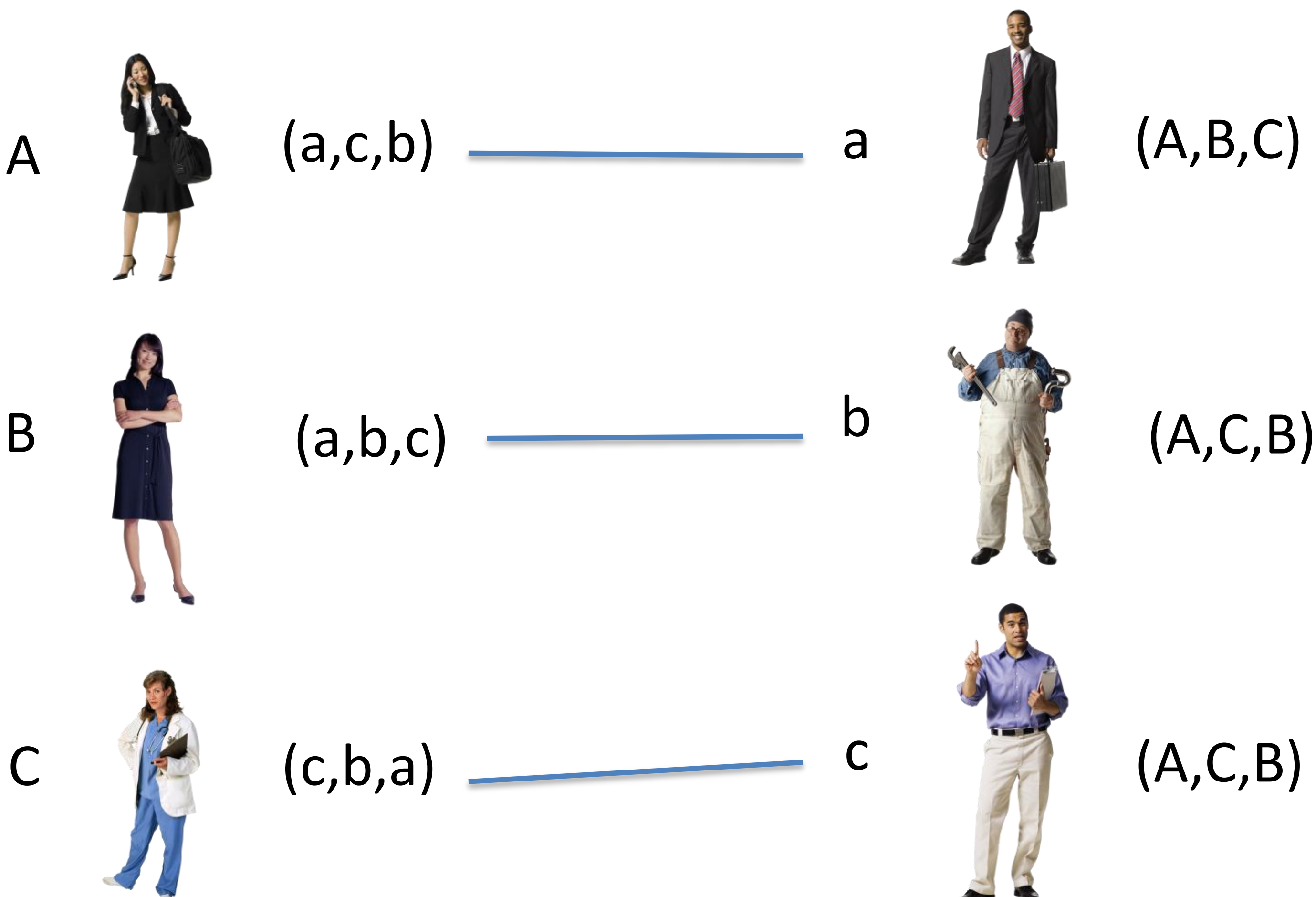
Matching

- Stable matching?



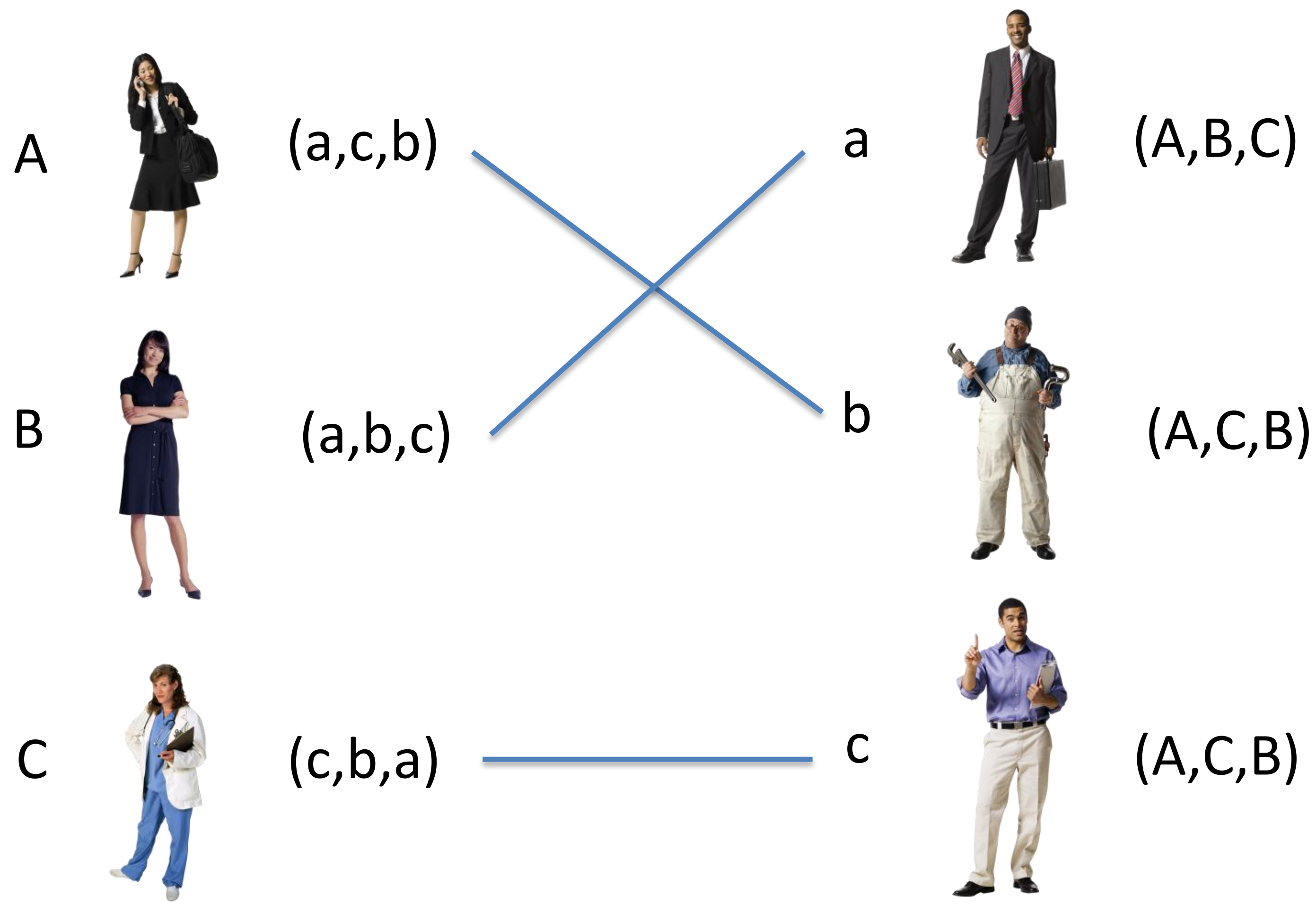
Matching

- Stable matching = no blocking pair



Unstable matching

- (A,a) is a blocking pair



Unstable matching

- (A,a) is a blocking pair

A  (a,c,b) — a  (A,B,C)

B  (a,b,c) — b  (A,C,B)

C  (c,b,a) — c  (A,C,B)

Existence of stable matching

- Gale-Shapley, 1962

Theorem* A stable matching always exists.

- Proof: construction of a stable matching

Centralized construction

- Step 1: each man proposes his favorite woman. Women accepts the best proposal (if several)

PROPOSALS

(a,c,b) A

a, b, c

a

(A,B,C)

(a,b,c) B

b

(A,C,B)

(c,b,a) C

c

(A,C,B)

Centralized construction

- Step 1: each man proposes his favorite woman. Women accepts the best proposal (if several)

CANDIDATE

(a,c,b) A  a

a  (A,B,C)

(a,b,c) B 

b  (A,C,B)

(c,b,a) C 

c  (A,C,B)

Centralized construction

- Step 2: rejected men propose their second choices.

CANDIDATE

(a,c,b) A  a

a  (A,B,C)

(a,b,c) B 

b  (A,C,B)

(c,b,a) C  b, c

c  (A,C,B)

Centralized construction

- Step 2: rejected men propose their second choices.

CANDIDATE

(a,c,b) A  a

a  (A,B,C)

(a,b,c) B 

b  (A,C,B)

(c,b,a) C  b

c  (A,C,B)

Centralized construction

- Step 3: rejected men propose their third choices.

CANDIDATE

(a,c,b) A  a

a  (A,B,C)

(a,b,c) B  c

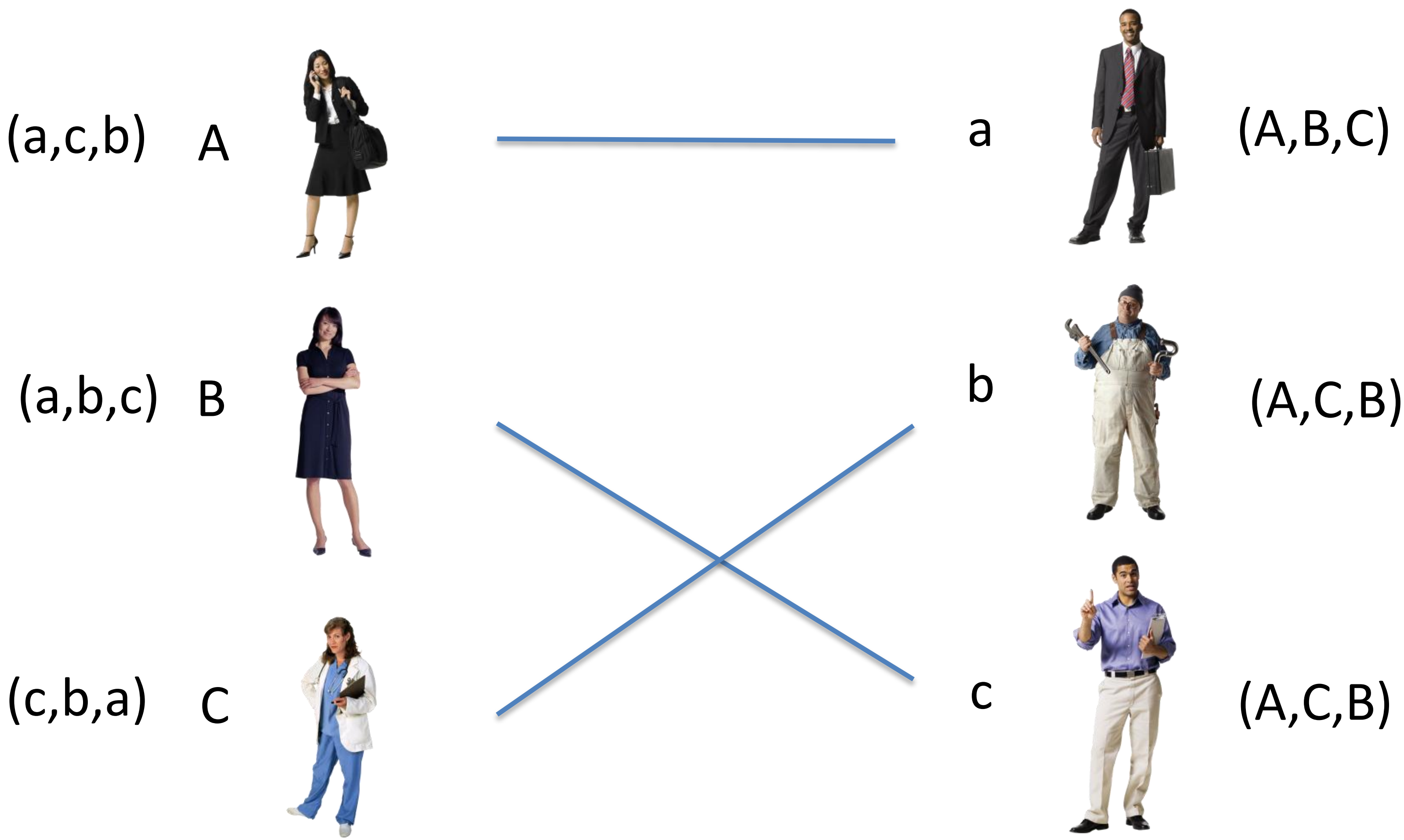
b  (A,C,B)

(c,b,a) C  b

c  (A,C,B)

Centralized construction

- Result:

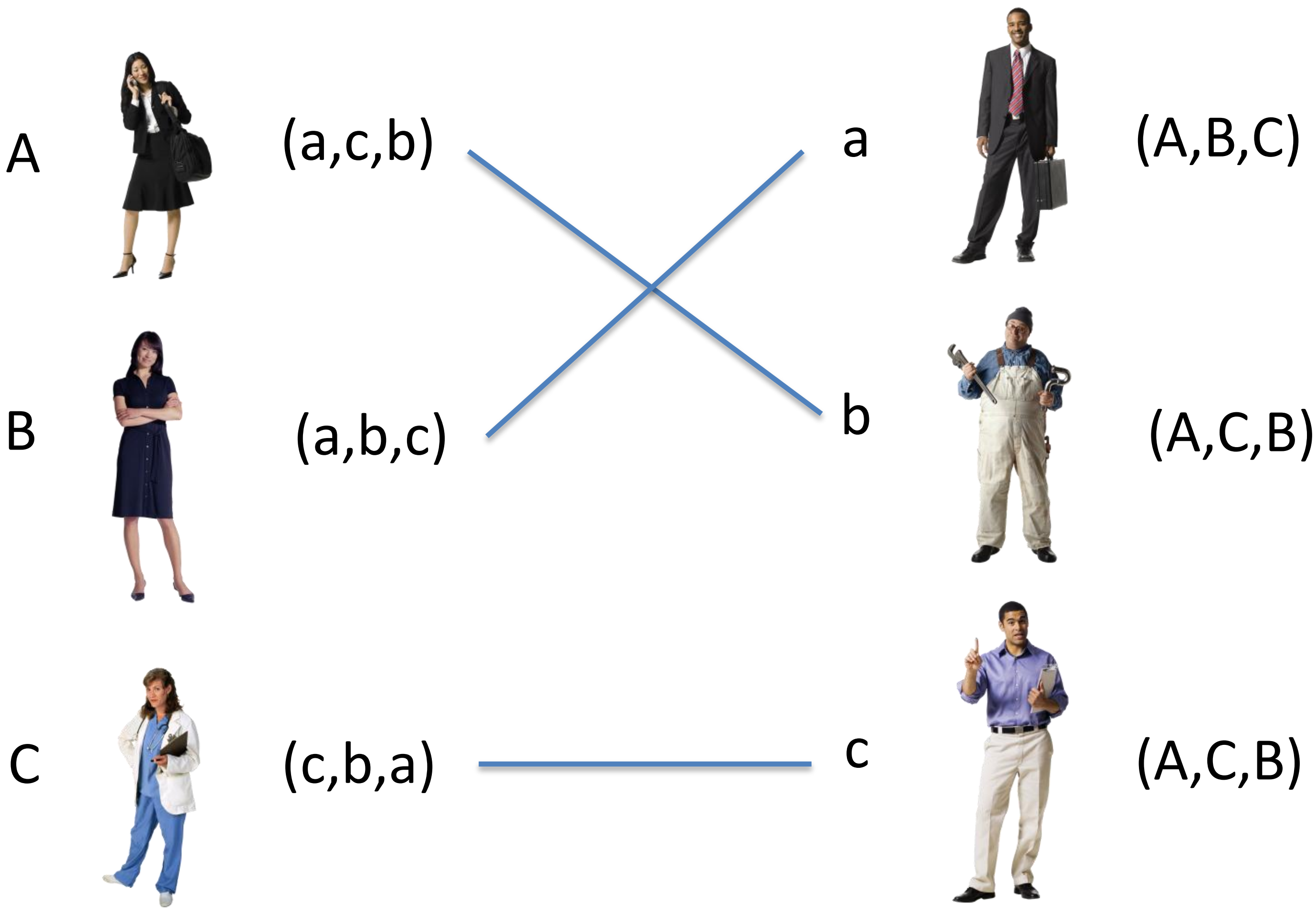


Complexity

- Gale-Shapley's algorithm finishes in at most $n^2 - 2n + 2$ steps
- A man proposes a given woman only once
- What about distributed algorithms?

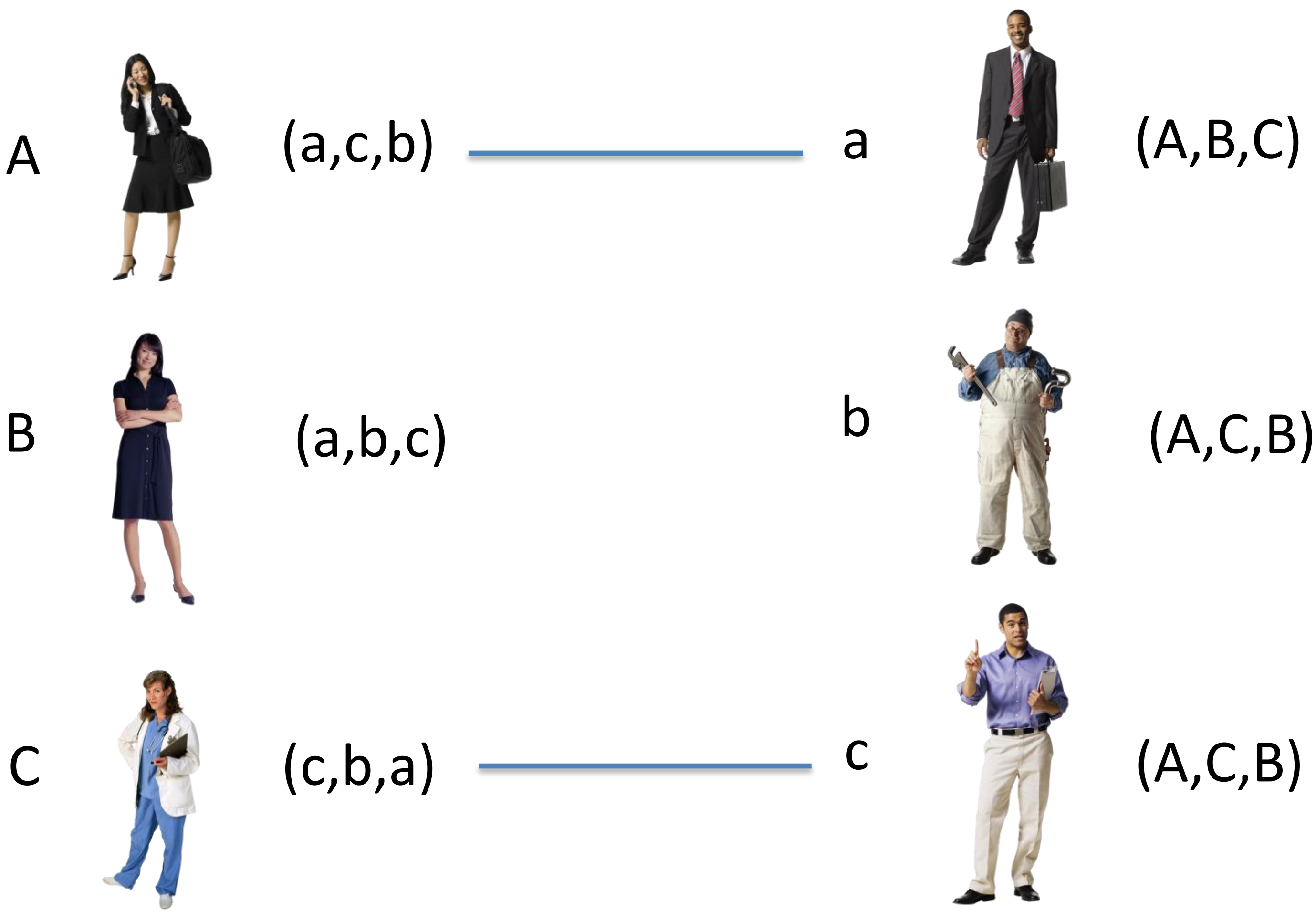
Best response dynamics

- Starting from any given unstable matching, a woman plays her best response (possibly breaking a marriage)



Best response dynamics

- Example: A proposes a, and wins him …



BR dynamics

- The best response dynamics can cycle (need 3 women and 3 men)*

- From every matching, there exists a sequence of BR of length $2n^2$ leading to a stable matching

- Random BR reaches a stable matching, but it can take an exponential time

* Uncoordinated two sided market, **Ackermann et al.**, EC, 2008

Outline of Part IV

- Games and equilibria
- Nash dynamics
- Fictitious play
- No-regret dynamics
- Trial and error learning

Fictitious play

Fictitious play

- Introduced by G. W. Brown 1951
- Principle:

  "Every player plays the best response action to the distribution of past actions of the other players."

Fictitious play

- Introduced by G. W. Brown 1951
- Principle: Bayesian interpretation

  “Every player assumes that each of the other players is using a stationary (i.e., time independent) mixed strategy. The players observe the actions taken in previous stages, update their beliefs about their opponents’ strategies, and choose the pure best responses against their beliefs.”

Discrete time fictitious play

- Empirical distribution of player-i's play up to time t:

$$p_i^t(s_i) = \frac{1}{t} \sum_{u=0}^{t-1} 1_{\{a_i^u = a_i\}}$$

- p^t: distribution on S given by the independent product of individual distributions p_i^t

- For stage t, player i selects action $a_i^t \in BR_i(p_{-i}^t)$

Continuous time fictitious play

- Empirical distribution of player-i's play up to time t:

$$p_i^t(s_i) = \frac{1}{t}\int_{u=0}^{t} 1_{\{a_i^u = a_i\}} du$$

- p^t: distribution on S given by the independent product of individual distributions p_i^t

- For stage t, player i selects action so that:

$$\frac{\partial p_i^t}{\partial t} \in BR_i(p_{-i}^t) - p_i^t$$

Discrete time: NE

Lemma If a pure strategy s is always played from a given time, then it is a pure NE.

Lemma If a pure NE is played at time t, then it is played thereafter.

Lemma If $\lim_{t \to \infty} p^t = p$, then the limiting distribution is a mixed NE.

Example of convergence

- Coordination game

		Player 2 a	Player 2 b
Player 1	A	(1,1)	(0,0)
	B	(0,0)	(1,1)

- There is convergence towards NEs

Example of non-convergence

- Shapley game

Player 2

		L	M	R
	T	(0,0)	(1,0)	(0,1)
Player 1	M	(0,1)	(0,0)	(1,0)
	B	(1,0)	(0,1)	(0,0)

- Cyclic behavior

Survey of existing convergence results

- Zero-sum 2x2 games: **Robinson**, 1951
- Super-modular games with unique equilibirum, **Milgrom-Roberts**, 1991
- 2xn games, **Berger**, 2003
- Super-modular games with diminishing returns, **Krishna**, 1992
- Weighted and ordinal potential games, **Monderer-Shapley**, 1996
- …etc.

Outline of Part IV

- Games and equilibria
- Nash dynamics
- Fictitious play
- No-regret dynamics
- Trial and error learning

No-regret

No-regret vs. Fictitious play

- Fictitious play: each player can observe the actions of other players, and compute best responses. Require the knowledge of the pay-off matrix of the game.

- No-regret: each player can observe her received pay-offs only. No need to know the number of players, the pay-off matrix.

An adversarial setting

- Idea: each player assumes that the other players' actions can be arbitrary, and try to do the best she can.
- The other players are replaced by an adversarial nature
- No-regret algorithms: an algorithm has zero regret, if asymptotically, after a sufficiently large number of stages, it performs almost *optimally*.

Lower bound on the regret

Theorem 2 For all K>1, for any time horizon T, there exists a distribution over pay-off assignments such that the regret of any online pay-off based algorithm is at least

$$\min(T, \sqrt{KT})/20$$

Exp3: a zero-regret algorithm

- Introduced by **Auer-Cesa Bianchi-Freud-Schapire**, 2002
- Algorithm:

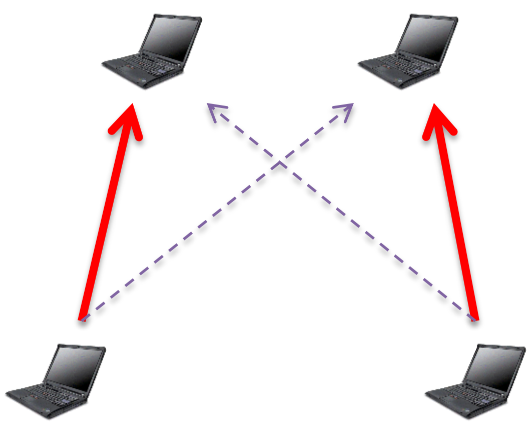
Parameter: $\gamma \in (0,1)$
Initialization: $w_a(1) = 1, \forall a \in A$
For each $t = 1, 2, \ldots$

1. Set
$$\forall a \in A, \quad p_a(t) = (1-\gamma)\frac{w_a(t)}{\sum_{a'} w_{a'}(t)} + \frac{\gamma}{K}$$
2. Draw a^t according to p^t
3. Receive pay-off $u^t(a^t) \in [0,1]$
4. For all $a \in A$, set
$$\hat{u}_a(t) = 1_{\{a=a^t\}} u^t(a)/p_a(t)$$
$$w_a(t+1) = w_a(t)\exp(\frac{\gamma \hat{u}_a(t)}{K})$$

Back to the game

- What if each player applies no-regret algorithms? Convergence to NEs?
- Know convergence results:
  - Convergence to NEs in constant-sum games, general sum 2x2 games, **Jafari-Greenwald-Gondek-Ercal**, 2001
  - Exp3 dynamics converge to weakly stable equilibria (efficient NEs) in congestion games, **Kleinberg-Piliouras-Tardos**, 2009
  - Extension of the previous results to the case of some ordinal potential games, **Kasbekar-Proutiere**, 2010
  - …etc.

Example: channel allocation



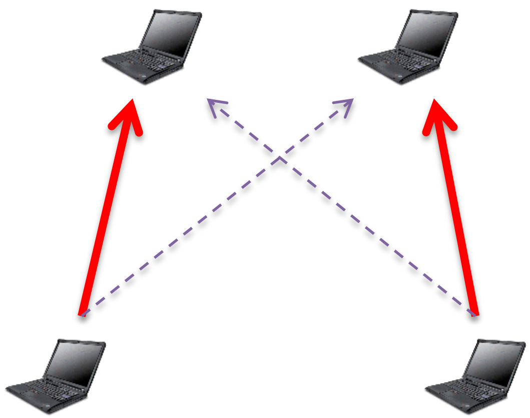
N links
m channels available for communication

Interaction through *interference*

Fading (unreliable transmissions)

Payoffs: link throughput (in bit/s)
(depends on interference and fading)

Interference

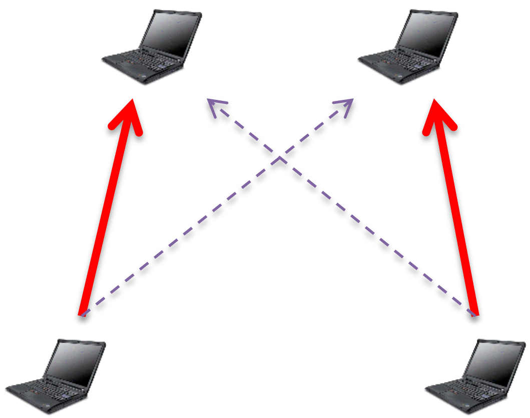


If two links simultaneously transmit on the same channel

- ***Collision***. None of the transmissions is successful

- ***Fair time sharing***. They share time fairly

Payoffs - Collisions



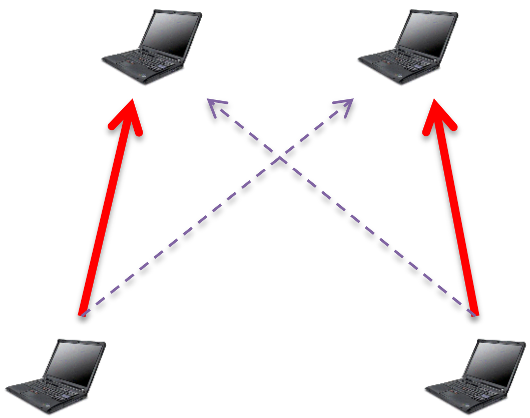
If link 1 transmits on channel *j* at time *t*, it receives a payoff R_1 equal to:

$$X_{1j} \times \underbrace{\prod_{i \neq 1} 1_{s_i(t) \neq j}}_{\text{interference}}$$

$X_{1j} \in \{0, 1\}$ random fading

$\mathbb{E}[X_{1j}] = \mu_{1j}$

Payoffs – Fair time sharing



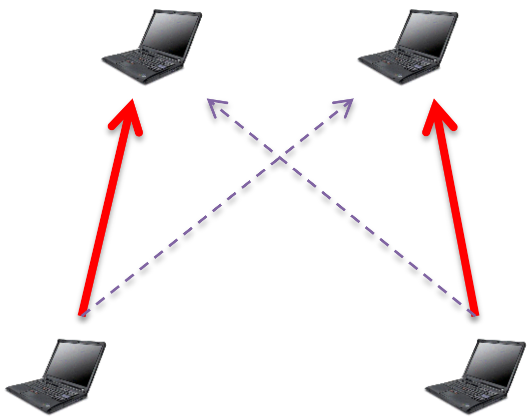
If link 1 transmits on channel *j* at time *t*, it receives a payoff R_1 equal to:

$$X_{1j} \times \underbrace{\frac{1}{|\{i : s_i(t) = j\}|}}_{\text{interference}}$$

$X_{1j} \in \{0, 1\}$ random fading

$\mathbb{E}[X_{1j}] = \mu_{1j}$

Constraints and Objective



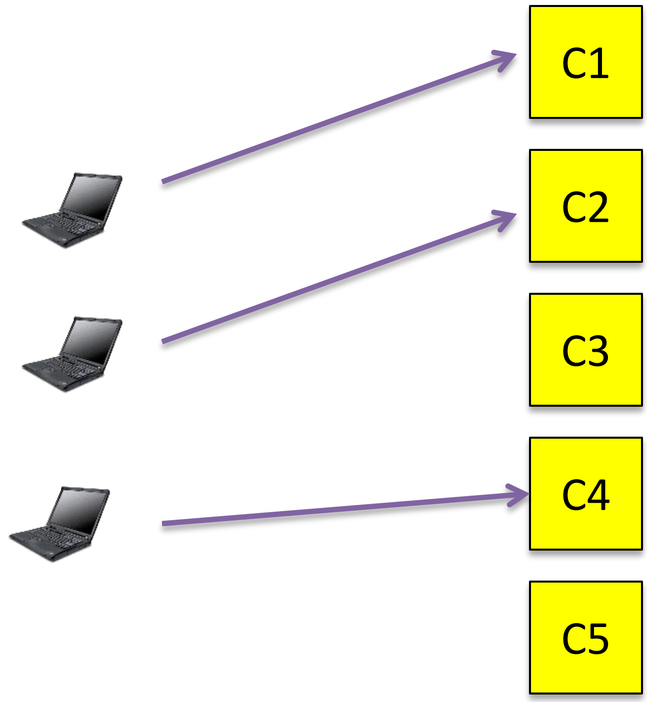
Lack of information

- Transmitter of link *i* has no a priori knowledge about channel conditions on her link
- Transmitter of link *i* has no a priori information about other links

Objectives: Transmitters should select channels so as to guarantee

- High network throughput
- Fairness

Multiple links



- i.i.d. sequences of payoffs: for all *i*

$$X_{ij}(t) \text{ i.i.d.} \qquad \mathbb{E}[X_{ij}(t)] = \mu_{ij}$$

- Each transmitter applies Exp3 to select a channel at each step, e.g. link 1 observes a payoff (collisions)

$$X_{1j} \times \prod_{i \neq 1} 1_{s_i(t) \neq j}$$

Result

Choose Exp3 parameter γ_t such that: $\sum_t \gamma_t = \infty, \sum_t \gamma_t^2 < \infty,$

e.g. $p_{ij}(t) = (1-\gamma_t)\frac{w_{ij}(t)}{\sum_l w_{il}(t)} + \frac{\gamma_t}{m}, \quad \forall j = 1, \ldots, m.$

Theorem Under Exp3, the system converges a.s. towards a pure Nash Equilibrium (one link per channel).

Proof

1. Stochastic approximation. The stochastic processes generated by Exp3 are asymptotic pseudo-trajectories of a system of ODEs

1. Analysis of the system of ODEs
    a. Fixed points (include all NEs)
    b. Convergence towards fixed points (Lyapounov analysis)
    c. Instability of fixed points that are not pure NEs

2. Exp3 stochastic processes cannot converge towards unstable fixed points

Step 1.

Theorem Almost surely,

$$\lim_{t\to\infty} \sup_{0\le h\le s} \|\underbrace{P(t+h)}_{\text{Exp3}} - \underbrace{p(t+h)}_{\text{ODE with } p(t)=P(t)}\| = 0$$

where $$\frac{dp_{ij}}{dt} = p_{ij}(f_{ij} - \sum_{l=1}^{m} p_{il} f_{il})$$

$$f_{ij} = \mathbb{E}[R_i | i \text{ selects } j]$$

Exp3 mimics the **replicator dynamics***!

* Sandholm; Maynard-Smith, …

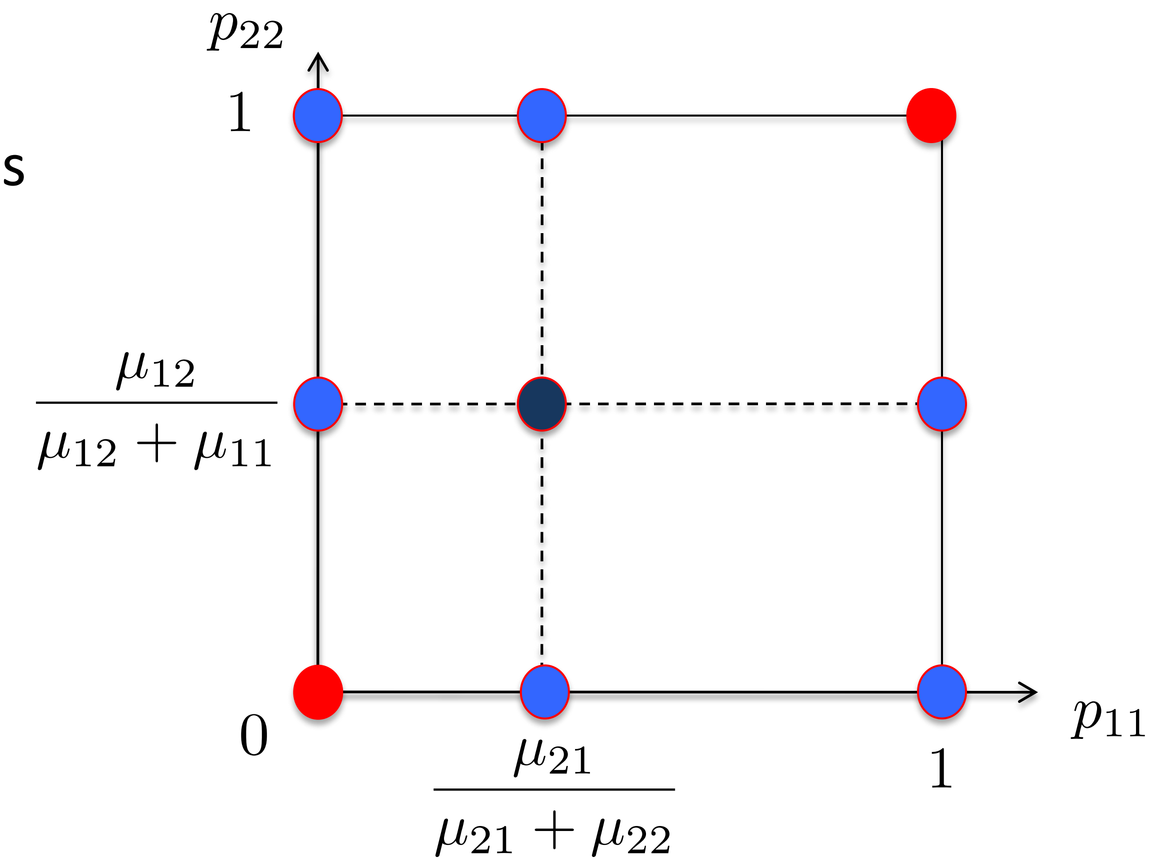
Step 2. Analysis of the ODE

Theorem All NEs are equilibrium points of the ODE. But There are many more fixed points.

2 users – 2 channels

Fixed points

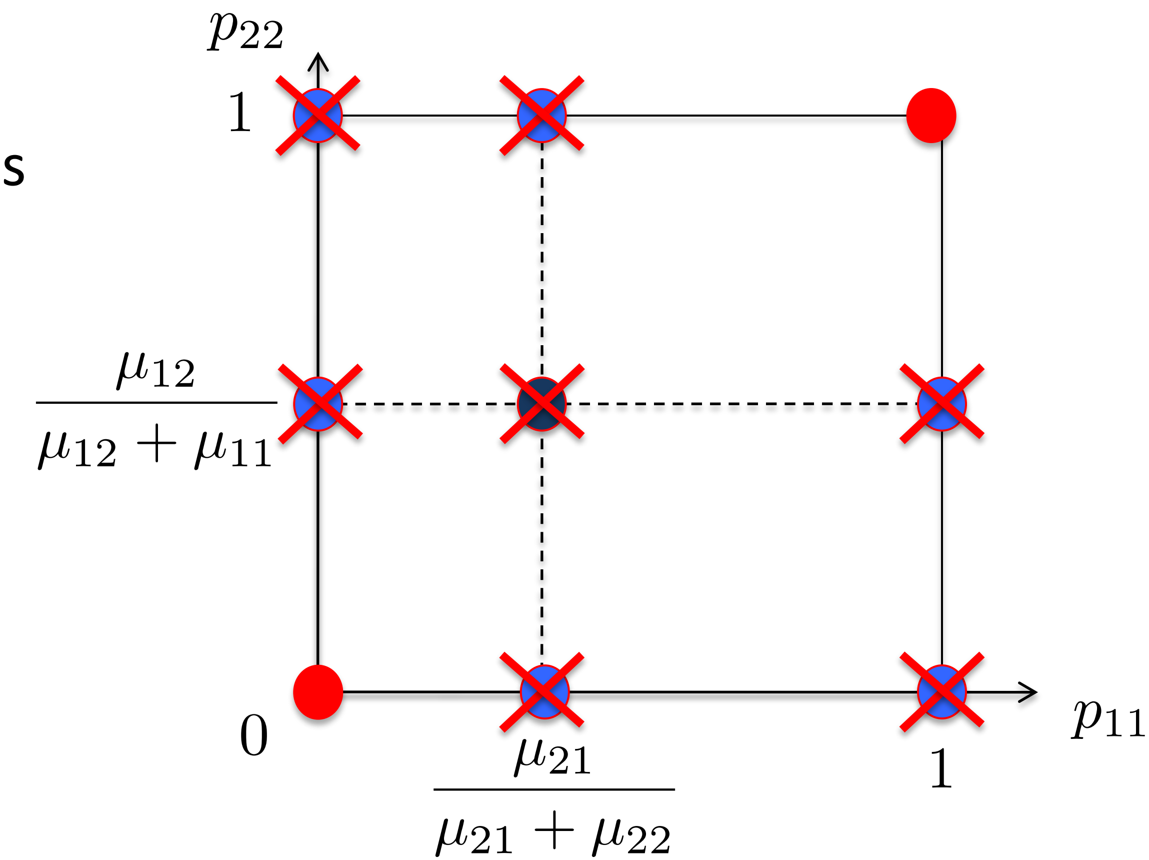
- Pure NEs
- Mixed NE
- Other



Step 2. Analysis of the ODE

Theorem Pure NEs are stable fixed points. The remaining fixed points are unstable

2 users – 2 channels



Fixed points

- Pure NEs
- Mixed NE
- Other

Step 2. Analysis of the ODE

Theorem From any initial condition, the ODE converges to a fixed point.

Step 3.

Theorem* Unlike the ODE, the stochastic process generated by Exp3 cannot converge to unstable fixed points.

* Pemantle, Annals of Probability 1990

2 links – 2 channels