Regret, stability & fairness in matching markets with bandit learners

Sarah Cen and Devavrat Shah

Department of Electrical Engineering & Computer Science, Massachusetts Institute of Technology.

Making informed decisions

Making an informed decision **requires knowledge** about options.
Ex. Choosing a career or housing.

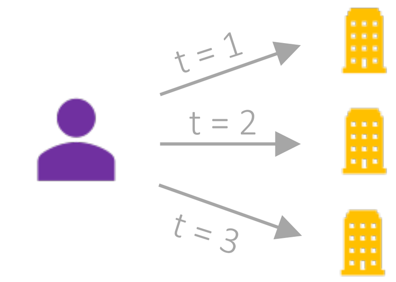
Learn thru **trial-and-error**
but not always possible **under competition.**

How does competition affect an individual's ability to make informed decisions and ultimately their long-term outcomes?

Our contributions

Learning +

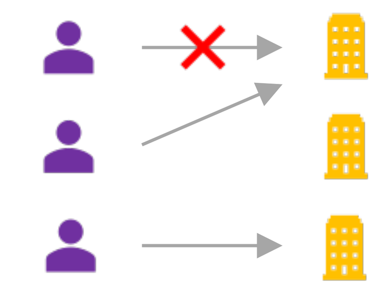
Bandits learn by sequentially sampling options.



Ex. Which careers are best for me? Learn through internships.

Competition =

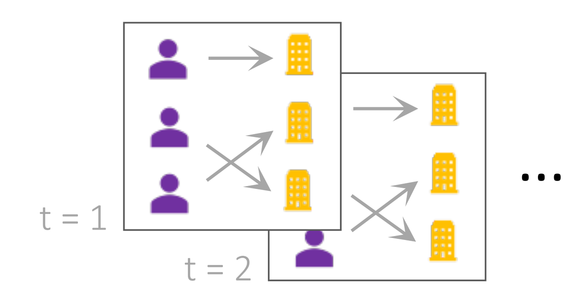
Agents compete for resources, a.k.a., matches.



Ex. Companies compete for workers, and vice versa.

Model ➡

Two-sided matching market with bandit learners



Ex. Learn career preferences while competing for internships.

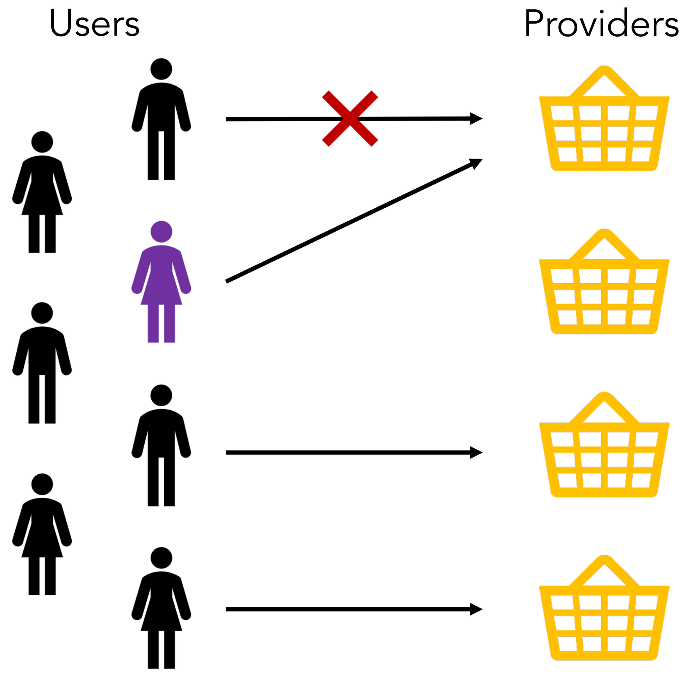
Main result

With costs and transfers, can simultaneously guarantee:

1. Stability ← good for platform
2. Low regret ← good for agents
3. Fairness ← good for society
4. High social welfare

Learning under competition

How to model? Combine game theory & RL: Matching + MAB.



Users have **preferences** over providers and vice versa.

Agents **compete** for matches.

Preferences are **unknown**.

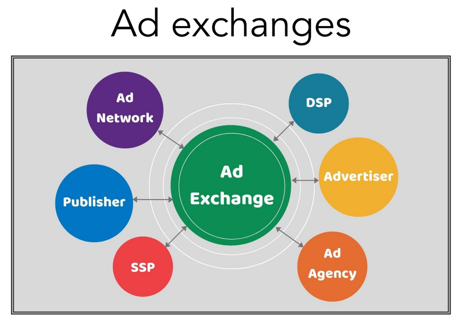
But agents **learn** them over time.

Challenge: How well do agents learn under competition?

Dating

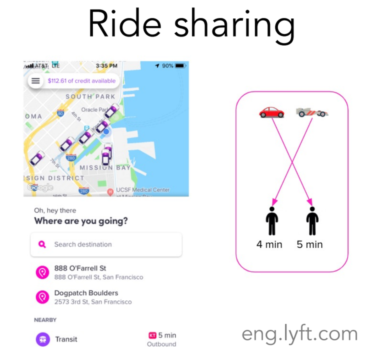
marketwatch.com

Ad exchanges



adpushup.com

Ride sharing



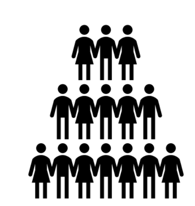
eng.lyft.com

Game theory:
- Competition.
- Preferences.
- Equilibrium.

Reinforcement learning:
- Learn thru interactions.
- Maximize reward.
- Explore vs. exploit.

Problem setup

N users



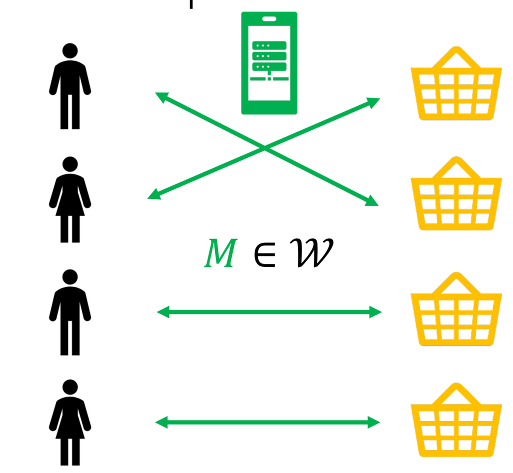
$\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$

$N \geq L$

L providers

$\mathcal{P} = \{p_1, p_2, \ldots, p_L\}$

Two-sided unknown preferences $\mu(a_1, a_2) \in \mathbb{R}$.



Centralized matching:

At $t = 0$, platform decides on $(\mathcal{M}, \mathcal{C}, \mathcal{T})$ → made known to all $\mathcal{A}$.

At each $t \in [T]$:

1. Update. Agents update estimates $\hat{\mu}_t$.
2. Report. Agents report UCB preferences ν_t
3. Match. Platform matches according to $\mathcal{M}(\cdot\,; \nu_t)$. Every agent a matched to a' receives sub-Gaussian rewards $X_t(a, a')$.
4. Pay and transfer. Every agent a matched to a' pays cost $\mathcal{C}(a, a'\,; \nu_t)$ and receives transfer $\mathcal{T}(a, a'\,; \nu_t)$.

Objectives:

Stability: No pair of agents is incentivized to defect.

Low (optimal) regret: Competition does not prevent learning.

Fairness: Regret is distributed evenly across agents.

High social welfare: Utilitarian measure of global performance.

Main results

Recent impossibility result [Liu et al. '20]:
Cannot simultaneously guarantee stable matching alongside low regret, fairness, and high social welfare.

We incorporate costs and transfers [Cen & Shah '22].
1. Model competition + exogeneous effects.
2. Can **guarantee stability, low regret, fairness, & high SW.**

Main theorems. Under mild conditions & balanced transfers, applying the Gale-Shapley algorithm at every time step ensures stability, fairness, and high social welfare. Moreover,

$$\underline{R}(a;\mathcal{M}) = \bar{R}(a;\mathcal{M}) = O\left(N^2 L\left(\frac{8\sigma^2\alpha\log T}{(\Delta^{\rho}_{\min})^2} + \frac{\alpha}{\alpha-2}\right)\right).$$

Moreover, there exists a pricing rule that simultaneously guarantees stability, fairness, and low regret.

Four proof ingredients:
1. GS algorithm at every time step → stability.
2. Costs & transfers must give unique true stable matching.
3. Ensure costs & transfers do not interfere with learning.
4. Cost & transfer rules do not require knowledge of μ.

LIDS LABORATORY FOR INFORMATION & DECISION SYSTEMS
MIT EECS
MIT