Learning Equilibria in Matching Markets from Bandit Feedback

Meena Jagadeesan*, Alexander Wei*, Yixin Wang, Michael I. Jordan, and Jacob Steinhardt (UC Berkeley)

arxiv.org/abs/2108.08843

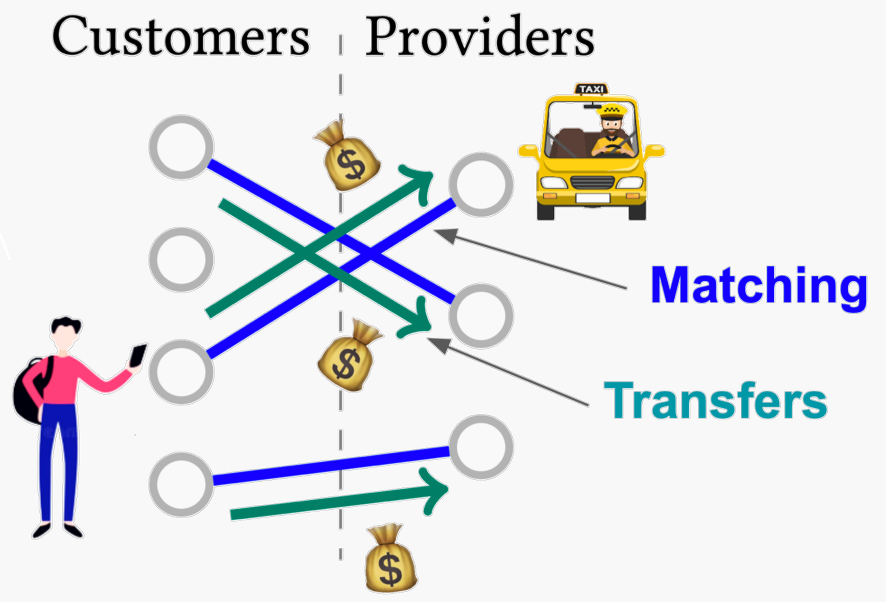
Berkeley UNIVERSITY OF CALIFORNIA

Our Contributions

1. Develop **bandit framework** for learning stable outcomes in matching markets
   - Capture learning in markets from noisy feedback
   - Introduce **Subset Instability** as a learning objective
2. Investigate algorithms for learning stable market outcomes
   - Design **no-regret algorithms** for the learning problem
   - Describe **preference structures** for which efficient learning is possible

Two-Sided Matching Markets

Uber airbnb DOORDASH upwork tinder THE MATCH NATIONAL RESIDENT MATCHING PROGRAM®

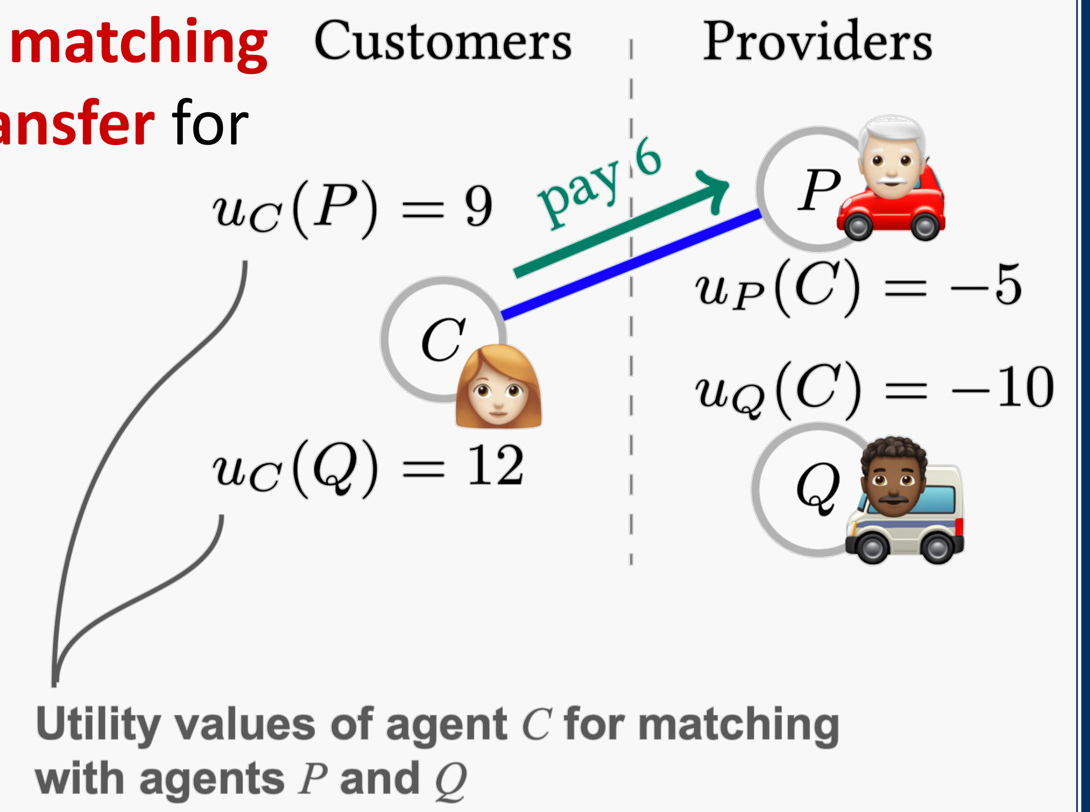


Matching Markets with Transferable Utilities

Platform selects **bipartite matching** along with a **monetary transfer** for each matched pair.

Incentive requirement = **stability**:

1. No "blocking" pairs
2. Individual rationality



A Framework for Learning Stable Matchings

Feedback Model

Matching + learning takes place over T rounds

In the t-th round:

- Agents $I^t \subseteq I, J^t \subseteq J$ arrive to the market
- Platform selects a matching with transfers (μ^t, τ^t)
- Platform observes noisy utilities $u_a(\mu^t(a)) + \varepsilon$ for each agent a

Platform incurs regret equal to **instability** of the selected outcome

Goal: Minimize **cumulative instability** over time

Subset Instability: An *Incentive-Aware* Loss Function

The **Subset Instability** of a market outcome (μ, τ) is defined to be:

$$\max_{S \subseteq I \cup J} \left[\left(\max_{\mu' \text{ over } S} \sum_{a \in S} u_a(\mu'(a)) \right) - \left(\sum_{a \in S} u_a(\mu(a)) + \tau_a \right) \right]$$

Interpretation:
Subset instability measures the maximum gain that any "coalition" S of agents could obtain by deviating from the given outcome (μ, τ) and only matching within S

Properties:

1. Subset Instability is 0 if and only if (μ, τ) is stable
2. Subset Instability $\geq$ the regret vs. welfare-maximizing matching
3. Subset Instability is equivalent to the "**minimum stabilizing subsidy**"
   - Shown via duality for an associated linear program

Algorithmic Results

A UCB-Based Algorithm

Theorem *(informal)***.** There exists an algorithm that incurs $\tilde{O}(N^{3/2}T^{1/2})$ instance-independent regret with N agents over T rounds.

Algorithm (MatchUCB):
Each round, select stable market outcome with respect to the **upper confidence bound estimates** of utilities.

This algorithm is **optimal** up to log factors!

Role of Preference Structure

For worst-case preferences, regret must scale *super-linearly* with the size of the market N.

When can we do better?

We explore two classes of preference structure:

- "Typed" preferences
- "Low-rank" linear preferences

Structure ⇒ can obtain $\propto N$ regret or better for each class

Extensions

1. $O(\log(T))$ **instance-independent** regret bounds
2. Interpretation of regret in terms of the platform's revenue
3. Extension of learning framework to **matching without transferable utilities** (the Gale-Shapley "stable marriage" setting)