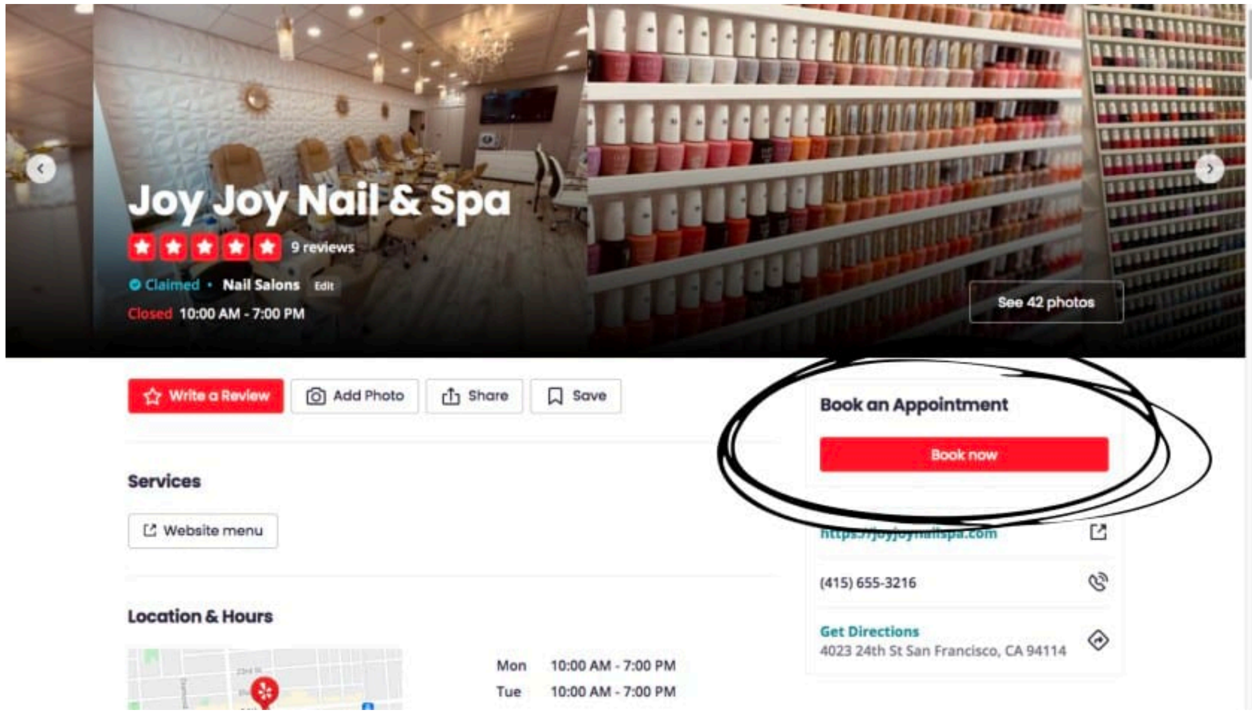
Interactive Learning with Pricing for Optimal and Stable Allocation in Markets

INFORMS 2022

Soham Phade
(Joint work with Efe Erginbas and Kannan Ramchandran)



Point of interest recommendations

E-commerce

Ride sharing and Delivery

Labor markets

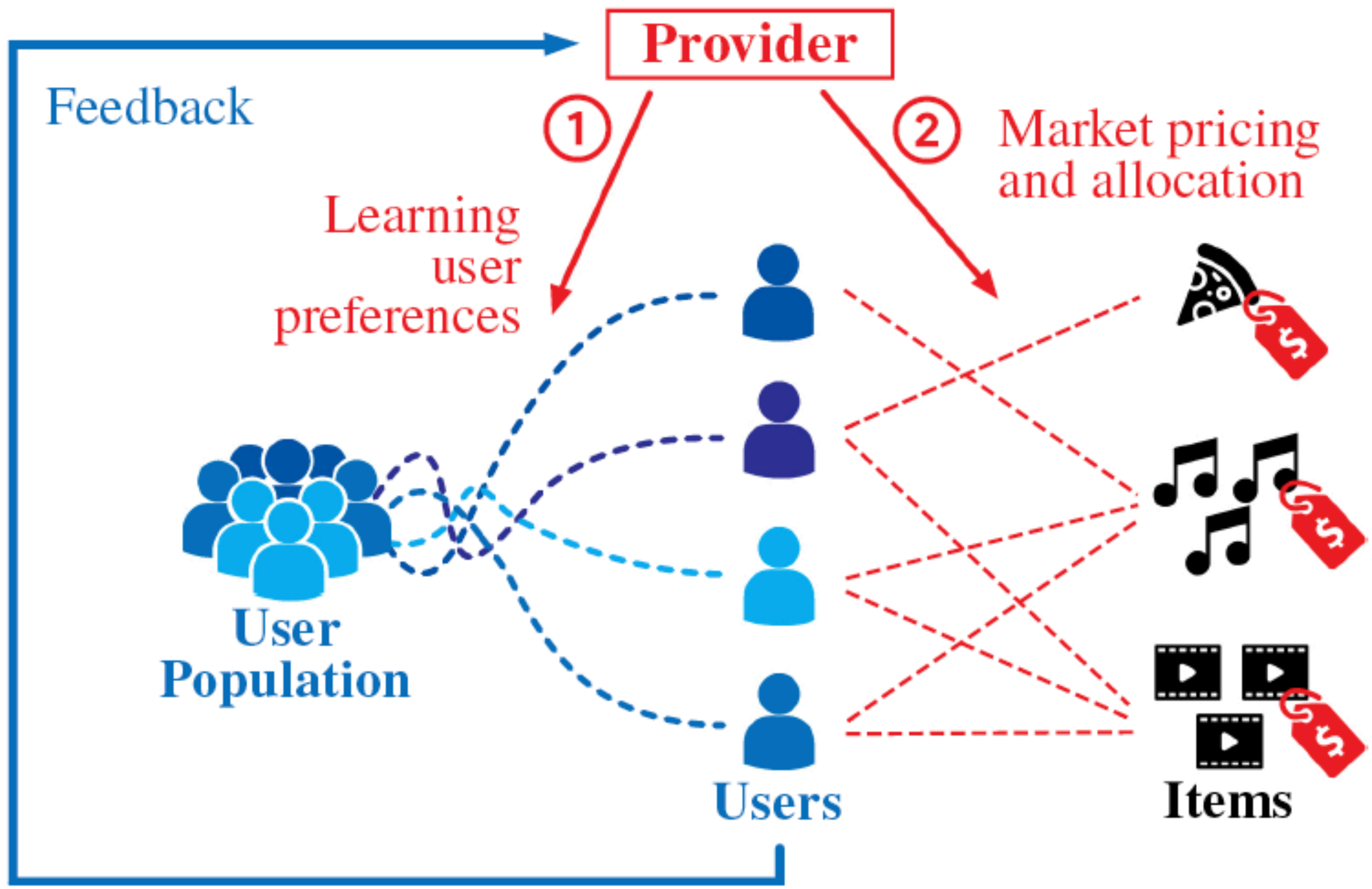
Main Challenges

- Large scale of operation
- User preferences unknown
- Learn user preferences and make recommendations
  - Exploit structure in preferences (eg. collaborative filtering)
  - Learn from interactive feedback (eg. multi-armed bandits, contextual bandits)
- Drawbacks:
  - Ignorant of capacity constraints
  - Results in overcrowding

Main Challenges

- Price discovery and allocation
  - Competitive equilibrium, Walras tatonnement process, dynamic pricing
  - Maximize social welfare
  - Envy-free and individually rational
- Drawbacks:
  - Assumes complete information
  - Assumes users can provide high dimensional responses

Market Aware Recommendation Systems



Our Approach

- Collaborative filtering: latent factor models
- Explore-exploit: OFU (optimism in face of uncertainty)
- Equilibrium pricing: Walrasian pricing

First to integrate all three aspects in one algorithm

What our Algorithm Achieves

- Has sub-linear social welfare regret across iterations
    - maximizing social welfare at each step is not possible since preferences are unknown
- Has sub-linear instability regret from user envy:
    - a user is said to have envy if she prefers a non-recommended item and measured by the difference in reward surplus when compared to the recommended item
- We provide theoretical guarantees

Setup

Modeling User Preferences

			
	0.8	0.6	0.2
	0.9	0.5	0.1
	0.2	0.3	0.6
	0.3	0.4	0.5

Items have limited capacities

User-Item Mean Reward Matrix Θ^*

Setup

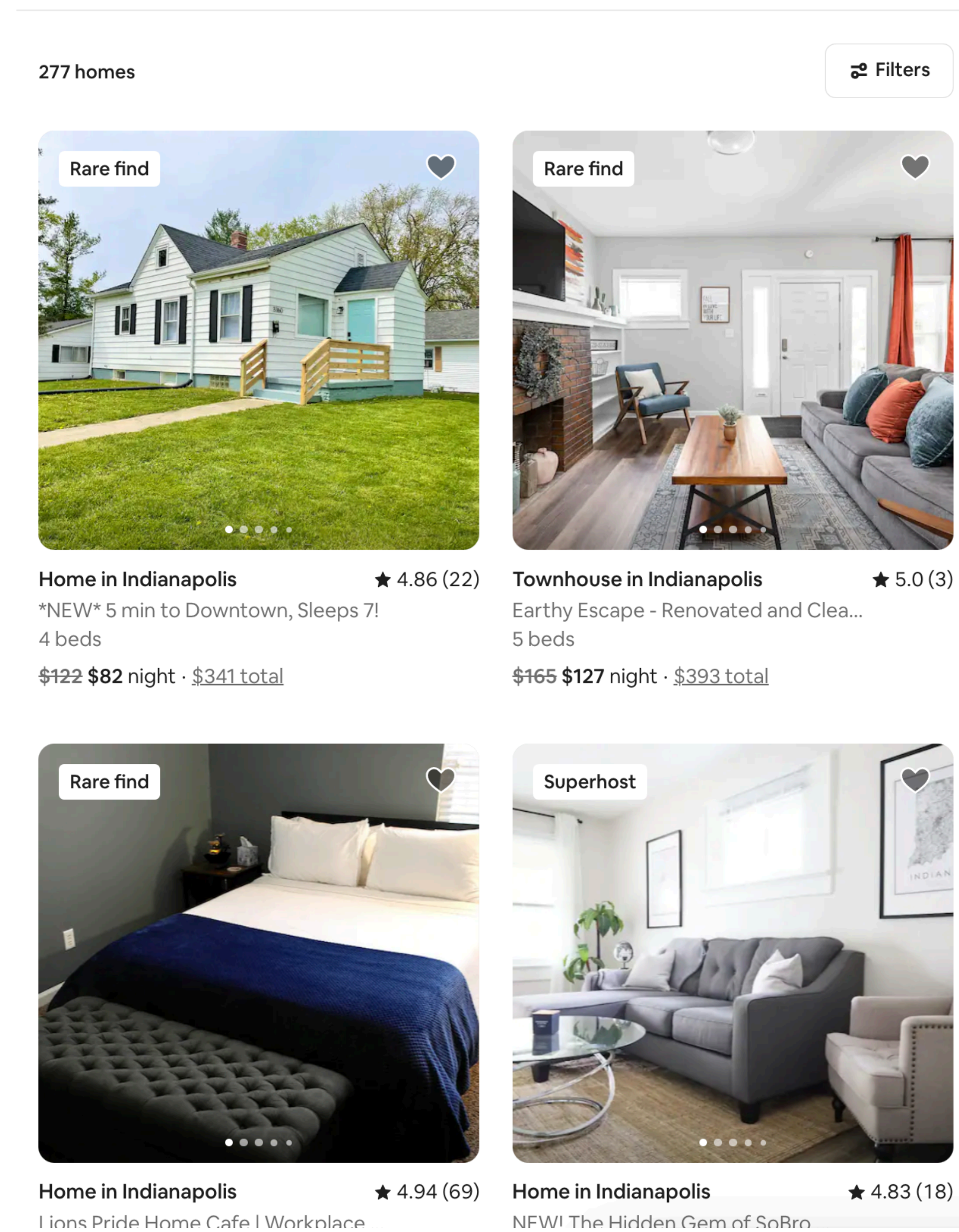
Interactive recommendation, allocation, and feedback

At each step a subset of users are active

active user

Recommendations

Select one and (Noisy) Reward Feedback



A Generic Algorithm

Interactive Learning for Allocation and Pricing (ILAP)

- Based on the collected information so far, find the least square estimate of the reward matrix under the structural conditions on preferences
- Consider confidence set around it with an appropriately defined metric and radius
- Optimistically solve the resource allocation problem with constraints assuming that the true rewards belong to this set
- Present the users with these allocations as recommendations at the corresponding shadow prices

Setting 1

Contextual Preferences

- Each item has a feature vector (known) (dim R)
- Each user has a feature vector (unknown) (dim R)
- A user-item reward is the linear product of these feature vectors
- These structural properties affect the first step in finding least squares estimate and the radius of confidence set
- Result: Avg. social welfare regret and instability regret of order

$$\tilde{O}\left(\frac{\sqrt{NMnR}}{\sqrt{T}}\right)$$

n max number of active users at any step
M number of items
N number of users
T step number

Setting 2

Low Rank Preferences

- We do not assume the item features to be known
- We assume the reward matrix to be of rank R
- Result: Avg. social welfare regret and instability regret of order

$$\tilde{O}\left(\frac{\sqrt{NM(N+M)R}}{\sqrt{T}}\right)$$

n max number of active users at any step
M number of items
N number of users
T step number

- If we do not assume any structure in preferences then we get

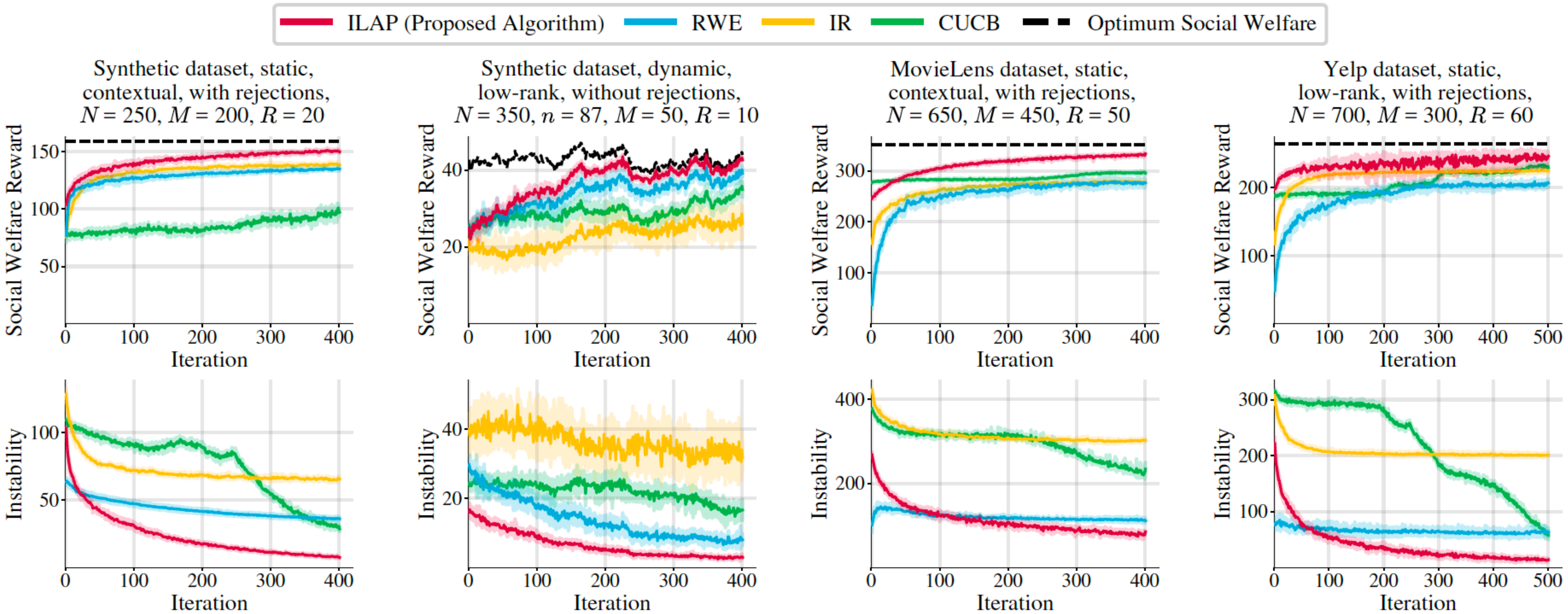
$$\tilde{O}\left(\frac{M\sqrt{Nn}}{\sqrt{T}}\right)$$

Giving user's an accept/reject choice

- Optimism in estimating preferences tends to raise prices
- Suppose a user accepts an item only if her reward is more that the offered price
- Then we have to lower the offered prices in proportion to the width of the confidence set
- This reduces the decay of regret in T to be

$$\tilde{O}\left(\frac{1}{T^{1/4}}\right)$$

Experiments



ILAP: Interactive Learning for Allocation and Pricing (Our Algorithm)
RWE: Recommendations without Exploration

IR: Interactive Recommendation
CUCB: Combinatorial UCB

Related Work

- Combinatorial multi-armed bandits: Audibert et al (2011), Chen et al (2013), Kveton et al (2015)
- Structured Linear Bandits: Combes et al (2017), Lu et al (2021)
- Bandits in economics: Liu et al (2020), Johari et al (2021), Jagadeesan (2021)
- Envy-free pricing: Guruswami et al (2005)
- Recommendation with capacity constraints: Christakopoulou (2017), Makhijani (2019)

Future Directions

- Show multiple recommendations at once instead of one
- Learn from user choice and not require user feedback
- Extending to multi-sided markets
- Lower bounds on regrets
- Maximizing revenue instead of social welfare

References

- Audibert, J.-Y., Bubeck, S., and Lugosi, G. (2011). Minimax policies for combinatorial prediction games. Proceedings of the Twenty-Fourth Annual Conference on Learning Theory, 19:107–132.
- Chen, W., Wang, Y., and Yuan, Y. (2013). Combinatorial multi-armed bandit: General framework and applications. Proceedings of the 30th International Conference on Machine Learning, 28(1):151–159.
- Kveton, B., Wen, Z., Ashkan, A., and Szepesvari, C. (2015). Tight Regret Bounds for Stochastic Combinatorial Semi-Bandits. Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, 38:535–543.
- Combes, R., Magureanu, S., and Proutiere, A. (2017). Minimal exploration in structured stochastic bandits. Advances in Neural Information Processing Systems, 30:1761–1769.
- Lu, Y., Meisami, A., and Tewari, A. (2021). Low-rank generalized linear bandit problems. Proceedings of the Twenty-Fourth International Conference on Artificial Intelligence and Statistics, 130:460–468.
- Liu, L., Mania, H., and Jordan, M. I. (2020). Competing bandits in matching markets. Proceedings of the Twenty-Third Conference on Artificial Intelligence and Statistics, pages 1618–1628.
- Johari, R., Manshadi, V., and Walton, N. (2021). Matching while learning. Operations Research.
- Jagadeesan, M., Wei, A., Wang, Y., Jordan, M. I., and Steinhardt, J. (2021). Learning equilibria in matching markets from bandit feedback. Advances in Neural Information Processing Systems.
- Guruswami, Venkatesan; Hartline, Jason D.; Karlin, Anna R.; Kempe, David; Kenyon, Claire; McSherry, Frank. (2005) On profit-maximizing envy-free pricing. Society for Industrial and Applied Mathematics. pp. 1164–1173. ISBN 978-0-89871-585-9.
- Christakopoulou, K., Kawale, J., and Banerjee, A. (2017). Recommendation with capacity constraints. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pages 1439–1448
- Makhijani, R., Chakrabarti, S., Struble, D., and Liu, Y. (2019). Lore: A large-scale offer recommendation engine with eligibility and capacity constraints. Proceedings of the Thirteenth ACM Conference on Recommender Systems, pages 160–168.