Combinatorial Pure Exploration with Limited Observation and Beyond

Yuko Kuroki

The University of Tokyo / RIKEN AIP

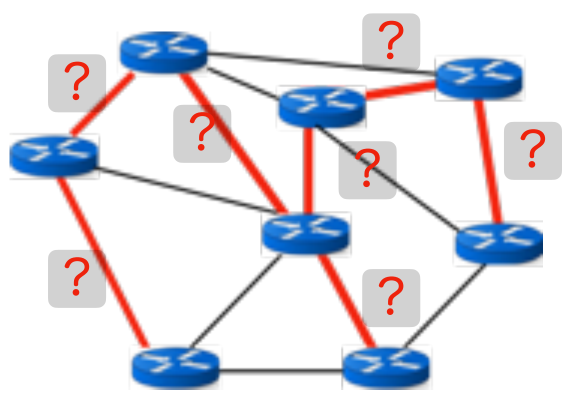
@2022 Data-driven Optimization Workshop, Online

2022/11/26

東京大学 THE UNIVERSITY OF TOKYO

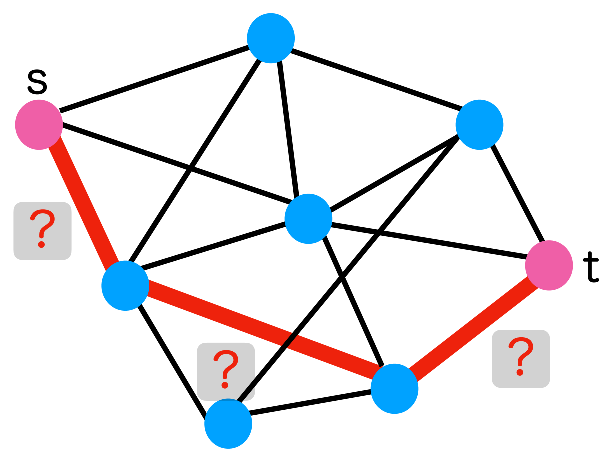
RIKEN

Decision Making with Combinatorial Actions



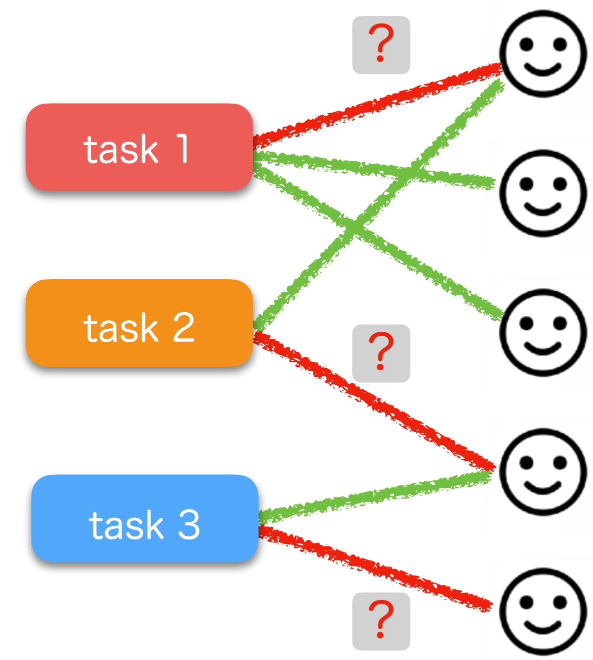
Spanning tree
in communication networks

Minimum spanning tree problem



s-t path
in road networks

Shortest path problem



Matching
from tasks to workers

Maximum weighted
matching problem

Input parameters might be
initially unknown or uncertain!

Input parameter must be learned over time!

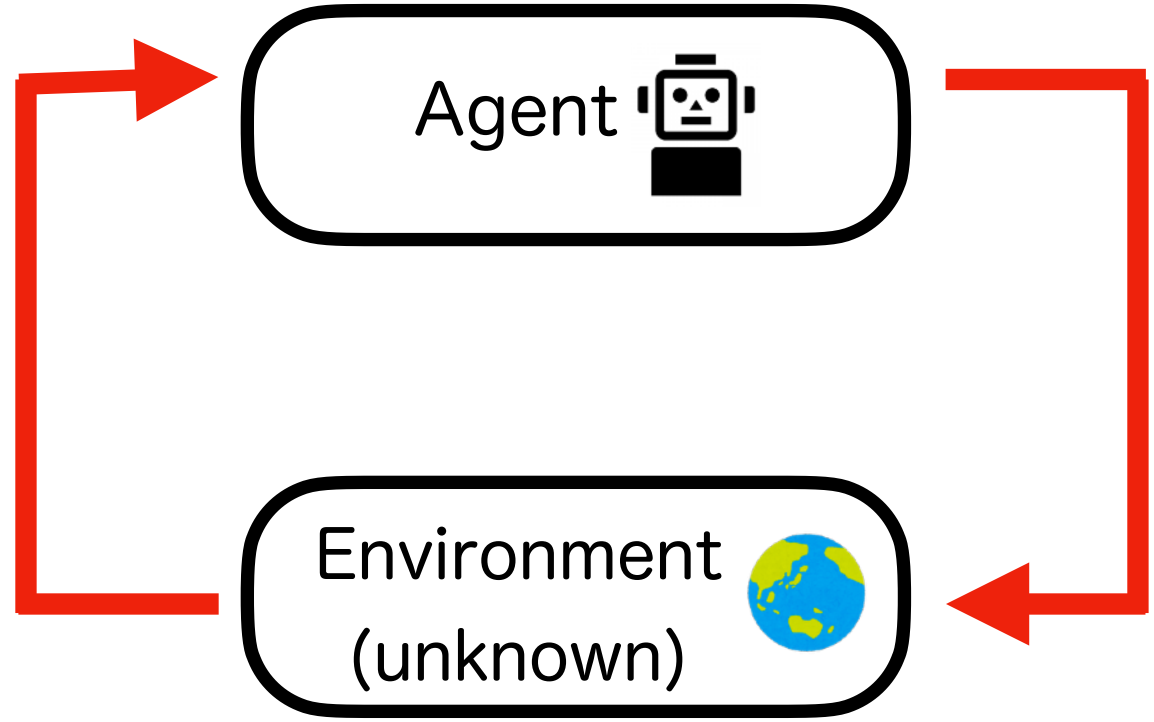
We focus on combinatorial bandits.

Combinatorial Bandits

[Cesa-Bianchi and Lugosi, 2006; Chen et al., 2013; Chen et al., 2014]

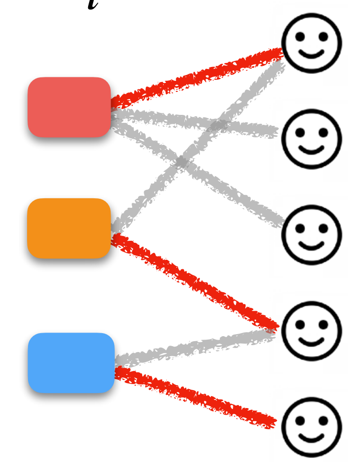
Observe (noisy) feedback

$r_{x_t} \in \mathbb{R}$



Pull a super arm

$x_t \in \mathcal{X}$



$[d] = \{1,2,\ldots,d\}$: a set of base arms (e.g., a set of edges)

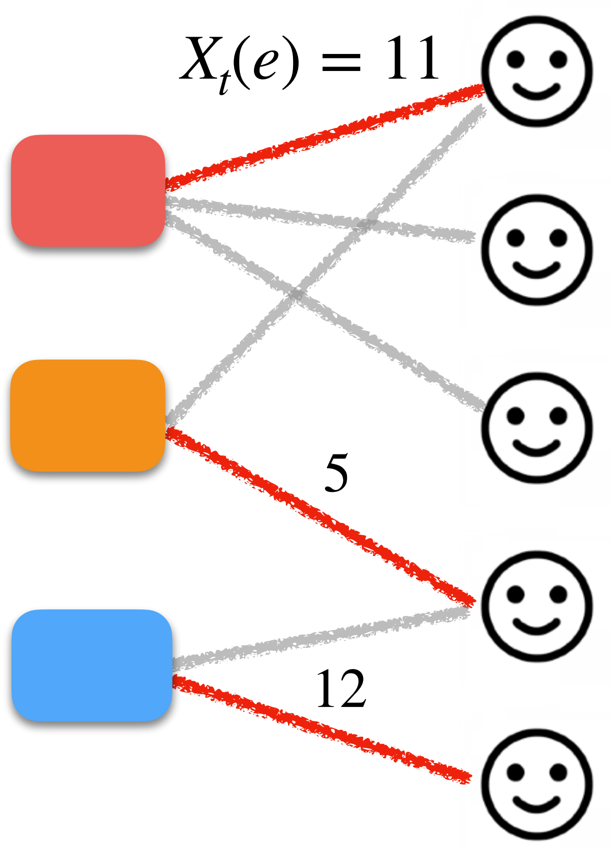
$\mathcal{X} \subseteq \{0,1\}^d$: combinatorial action space (e.g., spanning trees, paths, matchings)

$\theta \subseteq \mathbb{R}^d$: unknown parameters (e.g., edge weights)

Standard learning objectives

- Regret minimization: Minimize the cumulative regret
- Pure exploration: Identify the best super arm $x^* = \mathrm{argmax}_{x \in \mathcal{X}} x^\top \theta$ using as few exploration rounds as possible (This Talk)

Issue 1: Strong Observation



- Pull base arm $e \in [d]$ directly
  and observe $X_t(e) := \theta(e) + \eta_t$

Noise
R-subGaussian
(light tail)

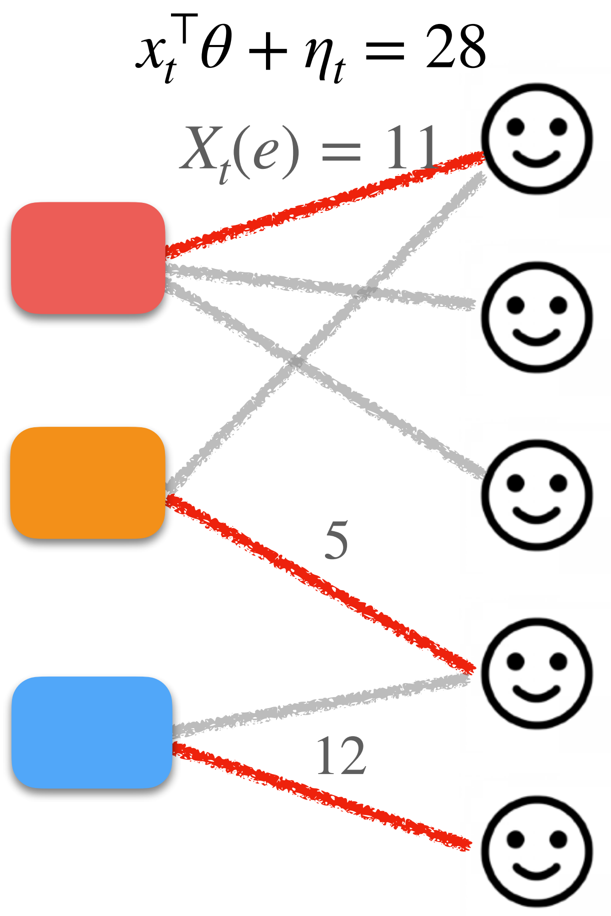
- Semi-bandit feedback:
  After sampling super arm $x_t \in \mathcal{X}$
  Observe all the elements in super arm

Issue 1

e.g. [Chen et al., 2014, 2016, Gabillon et al., 2016, Chen et al., 2017, Huang et al., 2018, Cao and Krishnamurthy, 2019; Joudan et al., 2021].

Due to practical constraints such as a budget ceiling or privacy concern, such strong feedback is not always available in recent applications.

Issue 2: Computational Issue



- **Full-bandit feedback (This study):**
  Pull super arm $x_t \in \mathcal{X}$,
  only observe sum of rewards $x_t^\top \theta + \eta_t$

- Linear reward case is a linear bandit
  →All existing algorithms for linear bandits need $O(|\mathcal{X}|)$ time complexity.

Issue 2

[Soare et al., 2014, Karnin, 2016, Tao et al., 2018, Xu et al., 2018, Zaki et al., 2019, Degenne et al., 2020, Katz-Samuels et al., 2020, Zaki et al., 2020, Jedra and Proutiere, 2020].

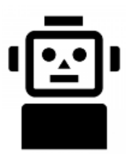
Since $|\mathcal{X}|$ is exponential size in d,
linear bandits algorithms cannot be applied to combinatorial setting.

Combinatorial Pure Exploration

$[d] = \{1,2,\ldots,d\}$: a set of base arms

$\mathcal{X} \subseteq \{0,1\}^d$: combinatorial action space (e.g., a family of indicator vectors of matchings, spanning trees, and paths)

$\theta \subseteq \mathbb{R}^d$: unknown latent vector



At round $t = 1,2,\ldots,T$

1. Choose super arm x_t (arm selection)
2. Observe random reward r_{x_t} (feedback)

$x^* = \mathrm{argmax}_{x \in \mathcal{X}} x^\top \theta$: maximum expected reward

$\mathtt{Out} \in \mathcal{X}$: output of the algorithm

Fixed confidence setting

Given confidence parameter $\delta \in (0,1)$, the agent must guarantee

$\Pr[\mathtt{Out} = x^*] \geq 1 - \delta$.

Evaluation metric : # samples the agent used to output (sample complexity)

Fixed budget setting

Given the sampling budget T, the agent minimizes the error probability

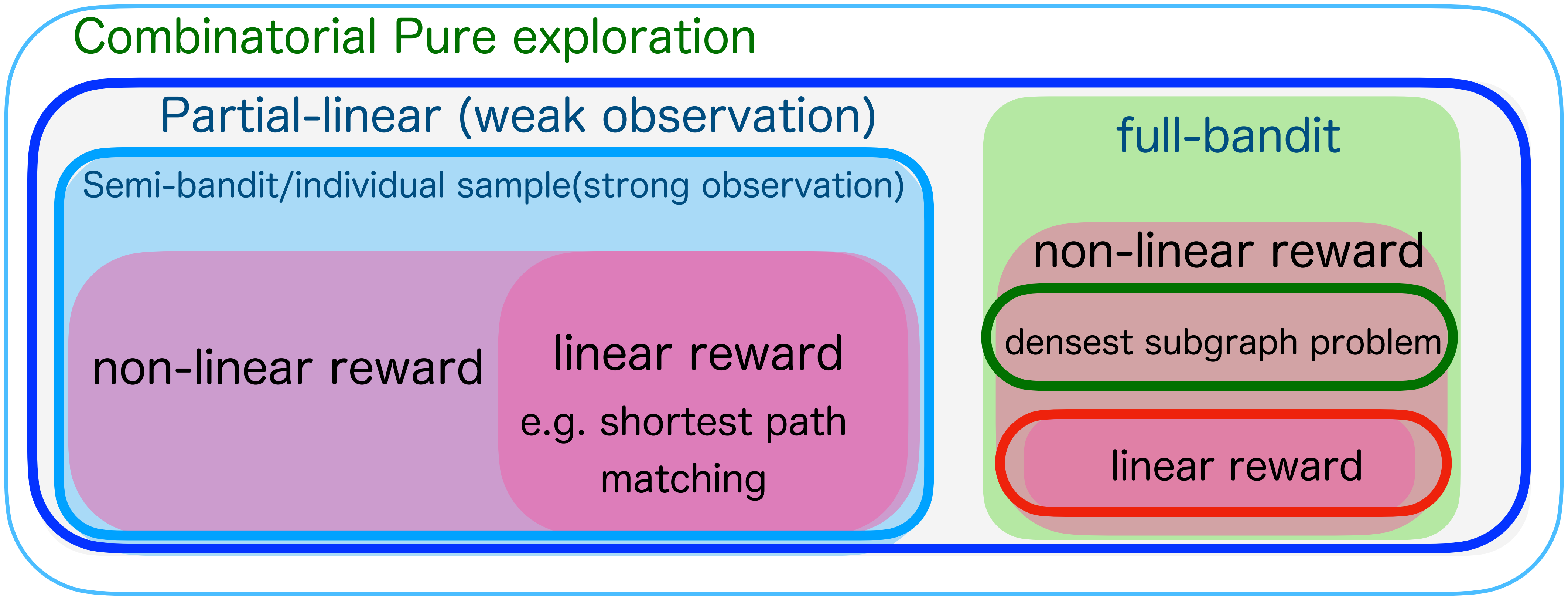
$\Pr[\mathtt{Out} \neq x^*]$

Evaluation metric : error probability $\Pr[\mathtt{Out} \neq x^*]$

Our Recent Advances

Existing study

individual sample/semi-bandit/linear reward/computational issue



Our study

- Kuroki+, Neco2020
- Kuroki+, ICML2020
- Du+, AAAI2021
- Du+, NeurIPS2021

- Y. Kuroki, L. Xu, A. Miyauchi, J. Honda, M. Sugiyama, Polynomial-time Algorithms for Multiple-Arm Identification with Full-bandit Feedback, Neural Computation, vol.32, no.8 pp.1733-1773, 2020.

This Talk: Limited Feedback

- Y. Kuroki, A. Miyauchi, J. Honda, M. Sugiyama, Online Dense Subgraph Discovery via Blurred-Graph Feedback, In Proc. International Conference on Machine Learning (ICML2020), pp. 5522-5532, 2020.
- Y. Du*, Y. Kuroki*, W. Chen, Combinatorial Pure Exploration with Partial or Full-Bandit Linear Feedback, In Proc. of Association for the Advancement of Artificial Intelligence (AAAI2021), 2021
- Y. Du, Y. Kuroki, W. Chen, Combinatorial Pure Exploration with Bottleneck Reward Function, In Proc. of NeurIPS 2021, 2021.
- Y. Kuroki, J. Honda, M. Sugiyama. Combinatorial Pure Exploration with Full-bandit Feedback and Beyond: Solving Combinatorial Optimization under Uncertainty with Limited Observation. (Preprint of the invited review article)

- Introduction
- Recent advances
  - Linear reward case with full-bandit feedback
  - Online densest subgraph discovery
  - Nonlinear reward and partial-linear feedback
- Open Problems
- Conclusion

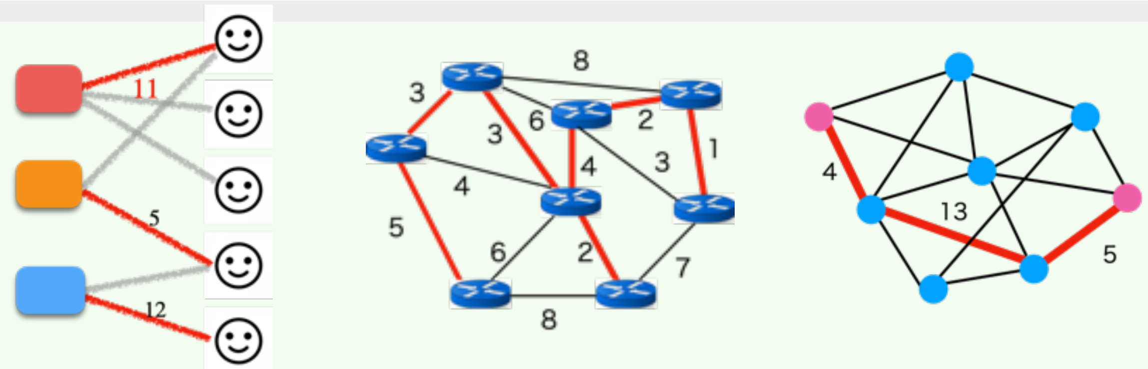
Problem setting for full-bandit linear case

$[d] = \{1,2,\ldots,d\}$: a set of base arms (e.g., a set of edges)

$\mathcal{X} \subseteq \{0,1\}^d$: combinatorial action space (e.g., spanning trees, paths, matchings)

$\theta \subseteq \mathbb{R}^d$: unknown parameters (e.g., edge weights)

- Reward function is linear $x^\top \theta$
- Full-bandit feedback, i.e., $r_{x_t} = x_t^\top \theta + \eta_t$ for chosen super-arm x_t



efficient oracle: $\mathrm{argmax}_{x \in \mathcal{X}} x^\top \theta$

$x^* = \mathrm{argmax}_{x \in \mathcal{X}} x^\top \theta$: optimal super arm with the highest expected reward

$\mathtt{Out} \in \mathcal{X}$: output of an algorithm

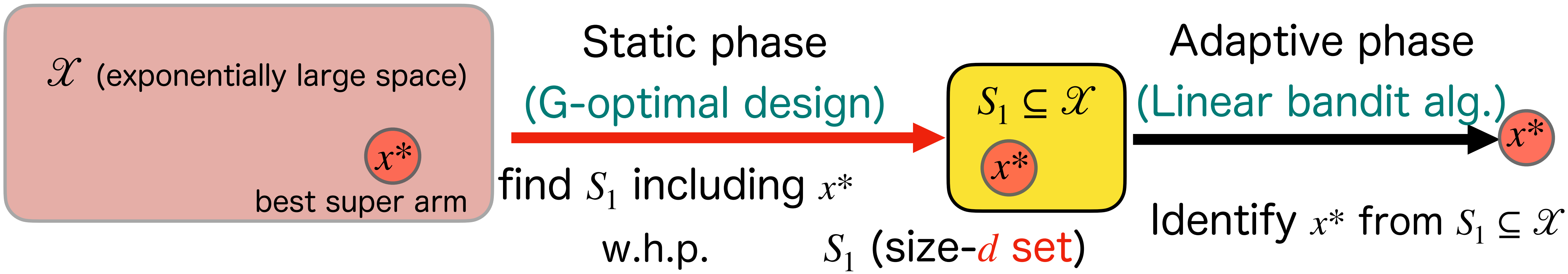
Fixed Confidence Setting

Setting: Given a confidence level $\delta \in (0,1)$, $\Pr[\mathtt{Out} = x^*] \geq 1 - \delta$ must be satisfied.

Evaluation metric: The number of samples used by an algorithm (i.e., sample complexity)

Sample Complexity

$\Delta_i\ (\geq \Delta_{\min})$: gap between the optimal super arm and i-th largest super arm



Main Theorem

Proposed algorithm guarantees $\Pr[\mathtt{Out} = x^*] \geq 1 - \delta$ and its sample complexity is:

$$T = O\left(\sum_{i=2}^{\left\lfloor \frac{d}{2} \right\rfloor} \frac{1}{\Delta_i^2} \left(\ln \frac{|\mathcal{X}|}{\delta} + \ln \ln \Delta_i^{-1} \right) + \frac{d(\alpha\sqrt{m} + \alpha^2)}{\Delta_{d+1}^2} \left(\ln \frac{|\mathcal{X}|}{\delta} + \ln \ln \Delta_{d+1}^{-1} \right) \right)$$

Adaptive phase — Static phase

where $\alpha = \sqrt{md/\xi_{\min}(\widetilde{M}(\lambda^*_{\mathcal{X}_\sigma}))}$ (approximation ratio of G-optimal design $\min_{\lambda \in \triangle(\mathcal{X})} \max_{x \in \mathcal{X}} x^\top M(\lambda)^{-1} x$)

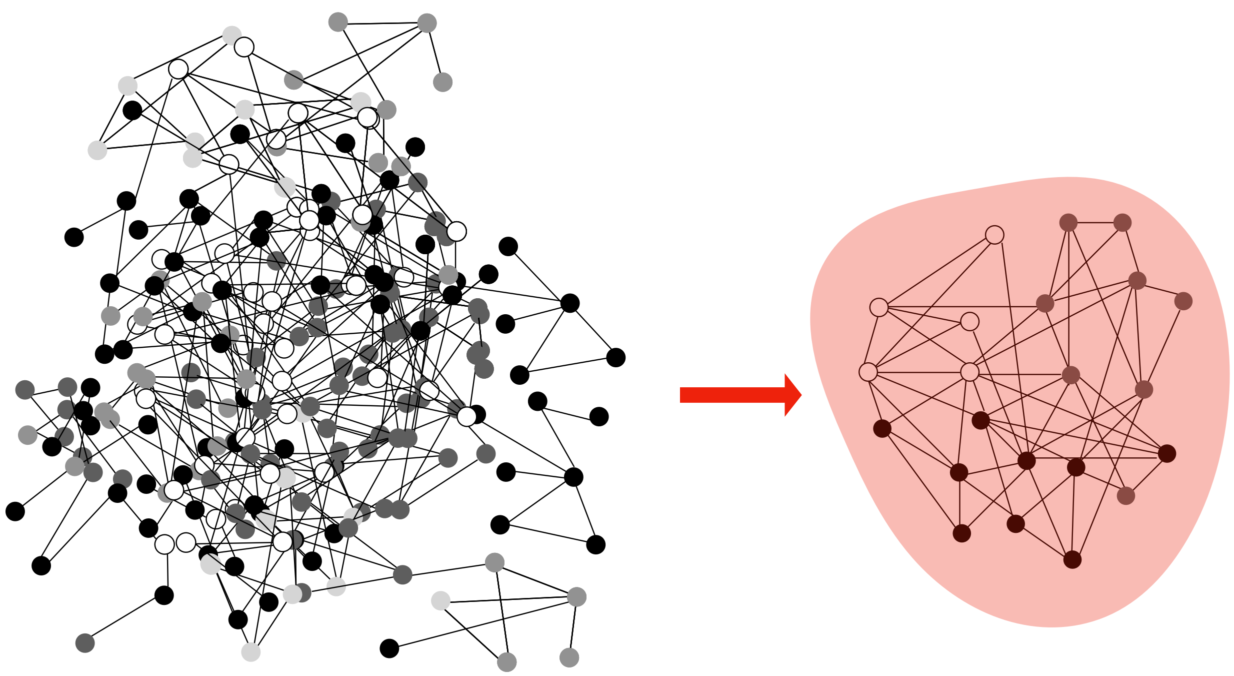
- This bound has mild dependence of $\Delta_{\min} (= \Delta_2)$
- It matches a lower bound for a family of instances (up to log factors)

• Y. Du*, Y. Kuroki*, W. Chen, Combinatorial Pure Exploration with Partial or Full-Bandit Linear Feedback, In Proc. of *Association for the Advancement of Artificial Intelligence (AAAI2021)*, 2021.

- Introduction
- Recent advances
  - Linear reward case with Full-bandit feedback
  - Online densest subgraph discovery
  - Nonlinear reward and partial-linear feedback
- Open Problems
- Conclusion

Dense Subgraph Discovery

Detecting dense components in networks is a fundamental task in graph mining

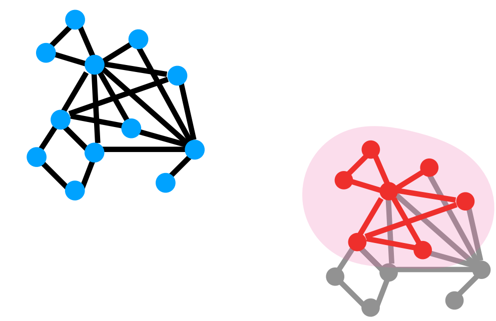


Applications examples

- Identifying molecular complexes in protein interaction networks
- Finding social groups in friendship networks
- Detecting communities and spam link farms in web graphs

Densest Subgraph Problem

Notation



- $G = (V, E, w)$: Edge-weighted undirected graph
- $E(S)$: a set of edges induced by a set of vertices S
- $w(S) = \sum_{e \in E(S)} w_e$: Sum of weights of the edges in S

Densest Subgraph Problem

Input: $G = (V, E, w)$ $(n = |V|$ & $m = |E|)$

Output: $S \subseteq V$ that maximizes $f(S) = \dfrac{w(S)}{|S|} \left(= \dfrac{\sum_{v \in S} \deg(v)}{2|S|} \right)$ (degree density)

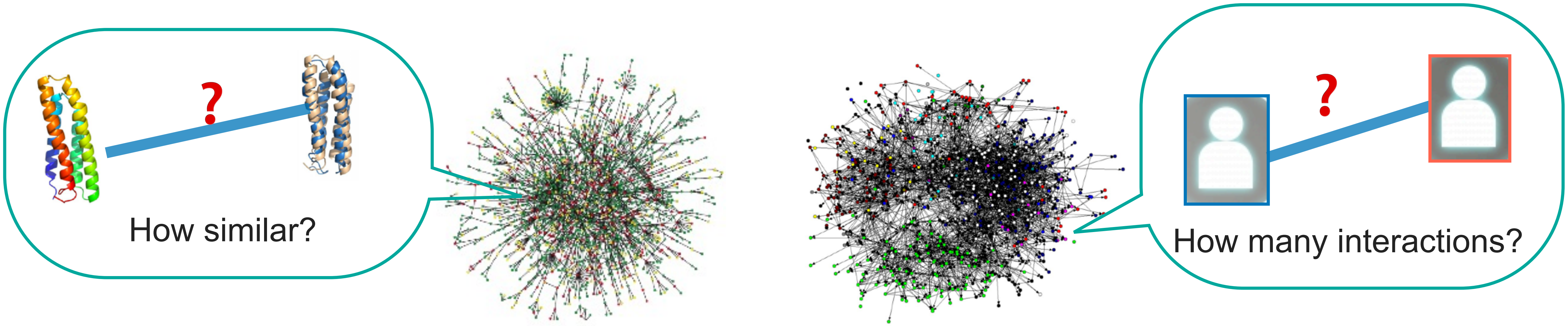
☺ Polynomial-time solvable! [Charikar'00;Goldberg'84]

Other problem variations

- Size-restricted variants [Andersen & Chellapilla'09, Feige et al. '01]
- Streaming settings [Angel et al.'12; Bahmani et al.'12; Bhattacharya et al. '15]
- Directed graphs [Charikar'00], Multi-layers graphs [Galimberti et al.' 17]
- Uncertain settings[Zou '13; Miyauchi & Takeda'18; Tsourakakis et al.'19]

Densest Subgraph with Uncertainty

- The graph data has uncertainty in real-world applications.



Protein-protein network

Email communication network

- How to handle the uncertainty of edge weights?

Existing model [Miyauchi & Takeda'18]

Robust optimization + Edge-sampling oracle

- All single edges are heavily and uniformly queried.
- It may be costly or may arise privacy concerns

Jeong et al. Nature 2001. 411 (3) and Rual et al. Nature 2005: 437 (4).
(picture) https://scx2.b-cdn.net/gfx/news/hires/2014/cellmembrane.jpg
https://spaceandorganisation.files.wordpress.com/2014/02/network_mediacompany_f2finteraction.jpg

Our Model: Bandit Formulation

[This Work]

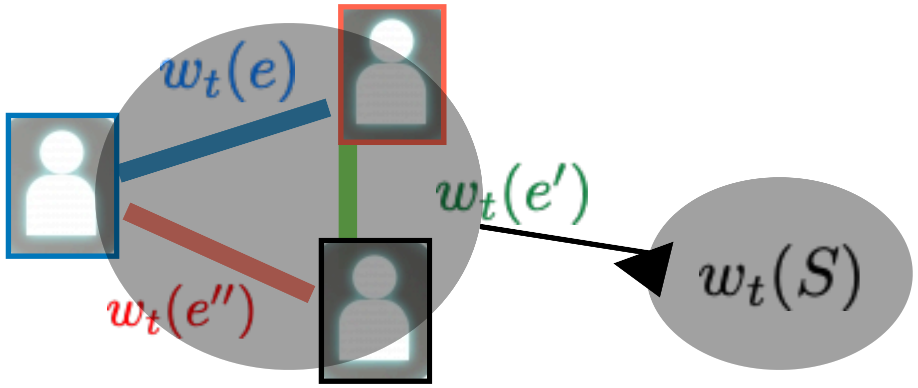
A novel learning framework for dense subgraph discovery by incorporating the concepts of multi-armed bandits

Our model

Pure exploration of multi-armed bandits

+ full-bandit feedback

- Sequentially observe a response from a set of edges
- Requires much less information of individuals



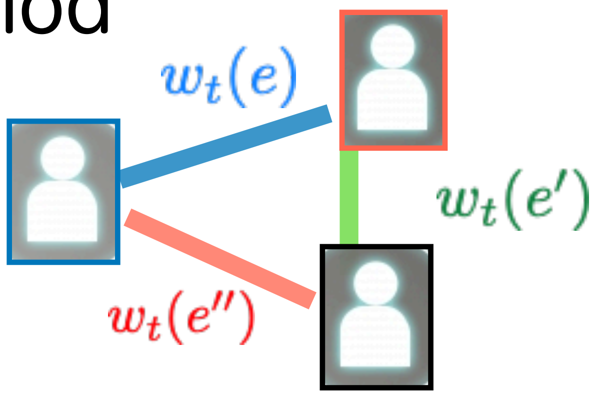
The total sum of the random weights in a queried subset can be observed

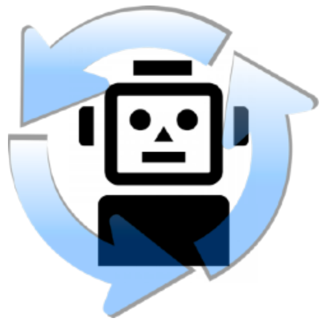
Problem Definition: Fixed Budget Setting

$w : E \to R_+$ is unknown to the agent

- At each round ($t = 1, \ldots, T$) in the exploration period
  - Chooses a set of edges E_t to sample
  - Observes the stochastic rewards $w^\top \chi_{E_t} + \eta_t$
  - Updates the sampling strategy

R-sub Gaussian





Problem (Densest subgraph in fixed budget setting)

Input: $G = (V, E, w)$ $(n = |V|$ & $m = |E|)$ and fixed budget T

Output: $S \subseteq V$ that maximizes reward function $f(S)$

Evaluation metric: the probability of error $\Pr[f(S_{Out}) \neq f(S^*)]$

*(approximate solution version $\Pr[f(S_{out}) < \alpha f(S^*)]$)

Proposed Algorithm

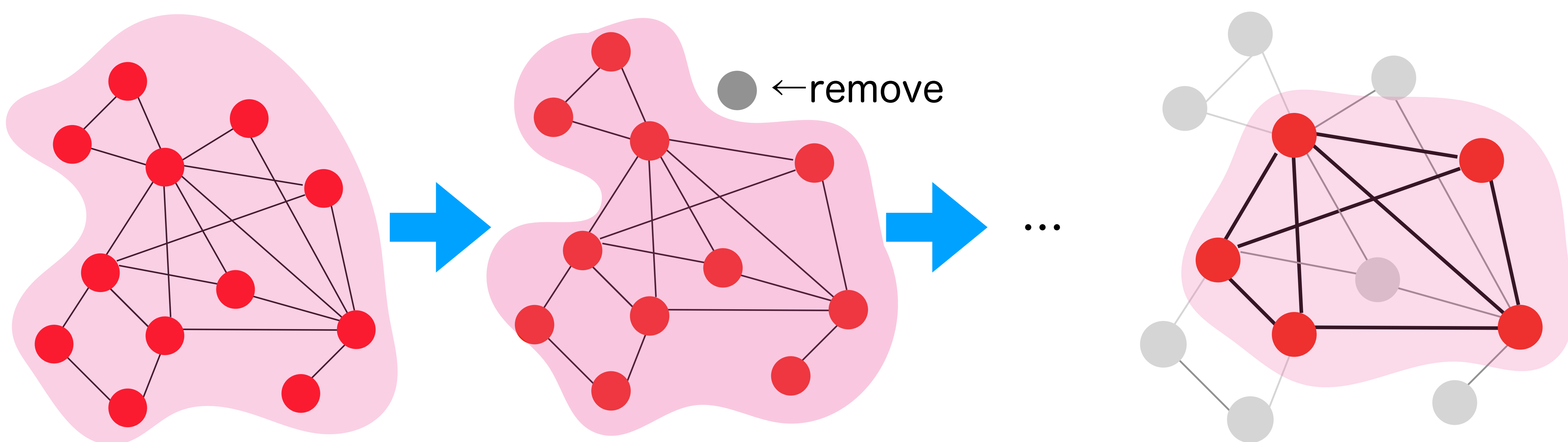
☺ Still be greedy in the face of uncertainty!

[Charikar, APRROX2000]

Greedy peeling: Almost linear time 0.5-approximation algorithm for the densest subgraph problem

[Audibert et al., COLT2010]

Successive reject strategy: One of the optimal sampling strategy for the BAI in multi-armed bandits



Theoretical guarantee

Main Theorem

- Given any $T > m$, and any latent edge weight w
- Assume that the edge weight distribution has R-sub-Gaussian tail.
- Then, Algorithm uses at most T samples and outputs a solution such that

$$\Pr\left[f_w(S_{\mathrm{OUT}}) < \frac{f_w(S^*)}{2} - \epsilon\right] \leq C_{G,\epsilon} \exp\left(-\frac{(T - \sum_{i=1}^{n+1} i)\epsilon^2}{4n^2 \mathrm{deg}_{\max} R^2 \tilde{\log}(n-1)}\right),$$

where $C_{G,\epsilon} = \frac{2\mathrm{deg}_{\max}(n+1)^3 2^n R^2}{\epsilon^2}$ and $\tilde{\log}(n-1) = \sum_{i=1}^{n-1} i^{-1}$.

- By setting the probability of error to a constant, the algorithm requires $T = \tilde{O}\left(\frac{n^3 \mathrm{deg}_{\max}}{\epsilon^2}\right)$ queries.
- We can guarantee the quality with polynomial-size samples!

• Y. Kuroki, A. Miyauchi, J. Honda, M. Sugiyama, Online Dense Subgraph Discovery via Blurred-Graph Feedback, In Proc. International Conference on Machine Learning (ICML2020), pp. 5522-5532, 2020.

Experimental Results for Dense Subgraph

Result: Performance of proposed algorithm in real-world graphs.

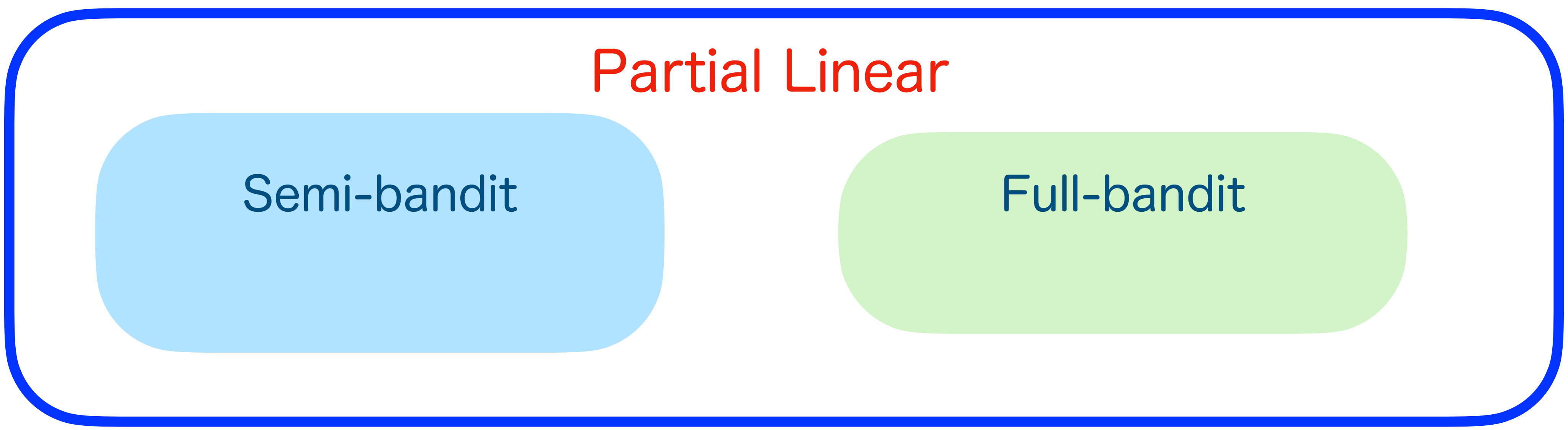
Graph	Proposed Algorithm				Robust-Sampling [Miyauchi &Takeda'18].			G-Oracle [Charikar'00]	OPT
	T	Quality	#Samples for single edges	Time(s)	Quality	#Samples for single edges	Time(s)		
Karate	10^3	111.08	58	0.00	111.08	10,296	0.02	111.08	111.08
Lesmis	10^4	177.66	752	0.02	179.72	51,816	0.07	176.29	179.72
Polbooks	10^4	227.43	419	0.02	228.67	214,767	0.22	227.47	228.67
Adjnoun	10^4	133.93	403	0.02	134.83	241,400	0.26	133.97	134.83
Jazz	10^5	599.42	6,837	0.4	599.43	1,115,994	1.49	599.43	599.43
Email	10^6	220.7	23,785	1.51	223.91	22,790,631	20.54	220.93	223.90
email-Eu-core	10^6	792.03	34,393	4.0	792.19	17,509,760	29.69	792.07	792.19
Polblogs	10^6	1211.37	16,508	4.38	1211.44	18,452,256	20.76	1211.44	1211.44
ego-Facebook	10^7	2654.40	103,546	42.61	2783.85	78,175,324	108.82	2654.44	2783.85
Wiki-Vote	10^8	1235.71	3,975,994	425.42	1235.95	288,205,696	638.92	1235.76	1235.95

Our algorithm significantly reduces the number of samples for single edges, compared to that of an existing state-of-the-art algorithm [Miyauchi & Takeda'18]

- Introduction
- Recent advances
  - Linear reward case with Full-bandit feedback
  - Online densest subgraph discovery
  - Nonlinear reward and partial-linear feedback
- Open Problems
- Conclusion

Partial Linear Feedback (CPE-PL)

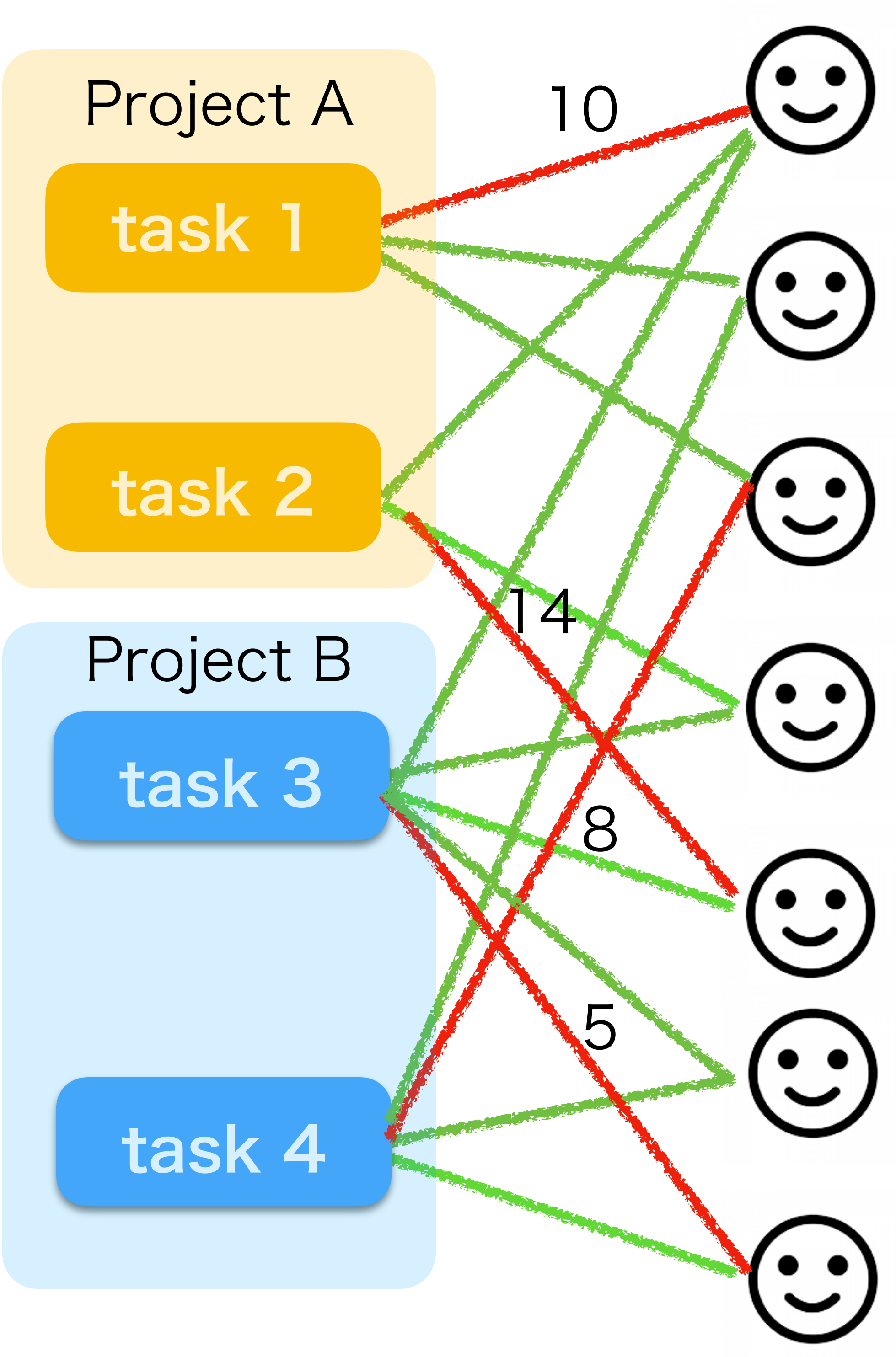
Can we go beyond the full-bandit?
Can we deal with nonlinear reward?



Lipschitz continuity

Assumtion 1. *There exists a constant L_p such that for any $x \in \mathcal{X}$ and any $\theta_1, \theta_2 \in \mathbb{R}^d$, $|\bar{r}(x, \theta_1) - \bar{r}(x, \theta_2)| \leq L_p ||\theta_1 - \theta_2||_2$.*

Applications Example of Partial-linear feedback



Semi-bandit : $\vec{y}_x = (10,14,8,5)$

Full-bandit : $\vec{y}_x = 37$

Partial-linear $\vec{y}_x = (24,13)$

Reward of Project A (10+14=24)

Reward of Project B (5+8=13)

Applications [Lin+, ICML2014]

- Online rankling with feedback from top-ranked items
- Task assignment in crowdsourcing with partial performance feedback

Sample Complexity

Lipschitz continuity

Assumtion 1. *There exists a constant* L_p *such that for any* $x \in \mathcal{X}$ *and any* $\theta_1, \theta_2 \in \mathbb{R}^d$, $|\bar{r}(x,\theta_1) - \bar{r}(x,\theta_2)| \le L_p ||\theta_1 - \theta_2||_2$.

Main Theorem

Proposed algorithm is δ-PAC and its sample complexity is:

$$T = O\left(\frac{|\sigma|\beta_\sigma^2 L_p^2}{\Delta_{\min}^2}\log\left(\frac{\beta_\sigma^2 L_p^2}{\Delta_{\min}^2\delta}\right)\right)$$

β_σ :upper bound of the estimate error σ :global observer set (support of pulls)

- General framework for nonlinear reward, limited feedback, and combinatorial structures.

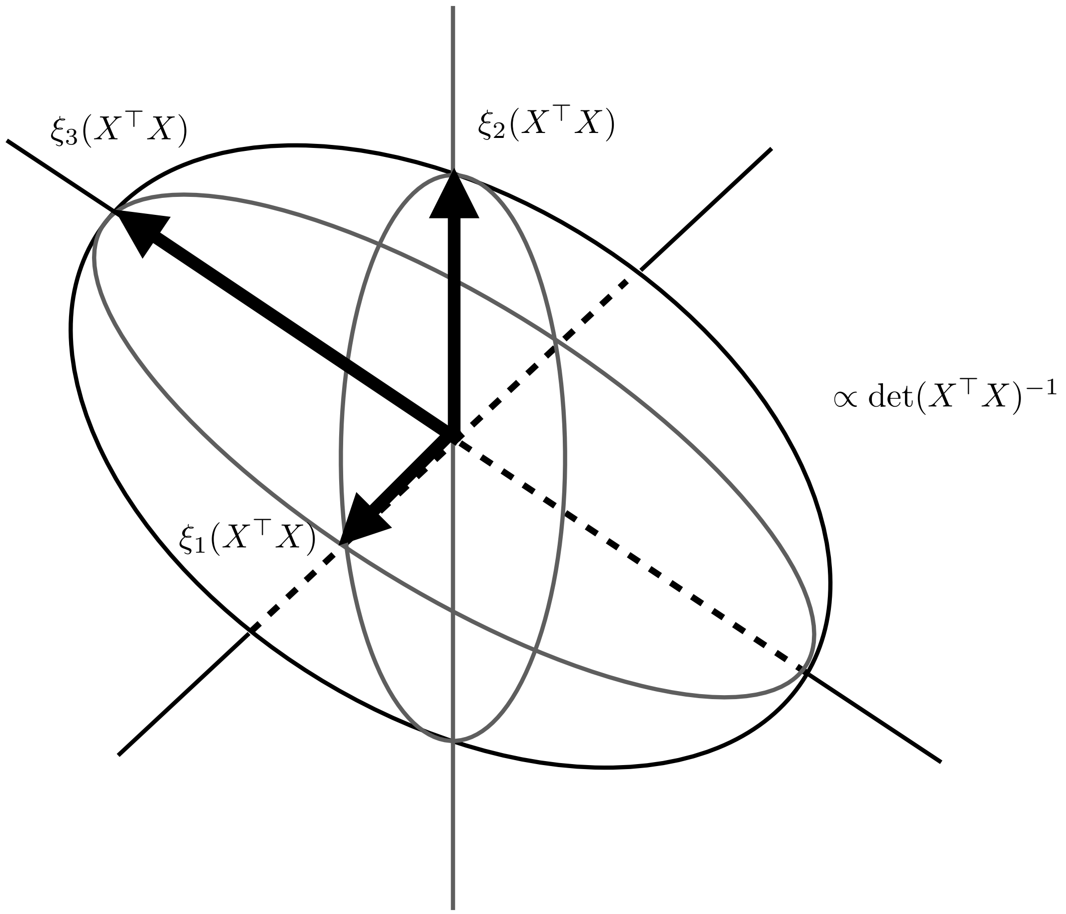
- The bound has heavy dependence on minimum gap
  →We need to design adaptive algorithms (Future work!)

- Y. Du*, Y. Kuroki*, W. Chen, Combinatorial Pure Exploration with Partial or Full-Bandit Linear Feedback, In Proc. of *Association for the Advancement of Artificial Intelligence (AAAI2021),* 2021.

- Introduction
- Recent advances
  - Linear reward case with Full-bandit feedback
  - Online densest subgraph discovery
  - Nonlinear reward and partial-linear feedback
- Open Problems and Conclusion

Open Problem 1: Experimental design

- G-optimal $\mathbf{x}_n^G = \mathrm{argmin}_{\mathbf{x}_n \in \mathbb{R}^{d\times n}} \max_{x\in\mathcal{X}} x^\top A_{\mathbf{x_n}}^{-1} x$
- E-optimal $\mathcal{X}_\sigma^* = \mathrm{argmin}_{\mathcal{X}_\sigma \subseteq \mathcal{X}} \lambda_{\max}((\sum_{x\in\mathcal{X}_\sigma} xx^\top)^{-1})$



Our study

- Naive approximation
- It results in worse sample complexity

Future work

- G-opt and E-opt is NP-hard
- Can we design good approximation algorithm?

Open Problem 2: Lower Bound

- It is open to prove a lower bound of polynomial-time δ-PAC algorithms, and design more efficient algorithms

Theorem for linear bandits [Fiez+. NeurIPS2019]

Any δ-PAC algorithms has sample complexity of

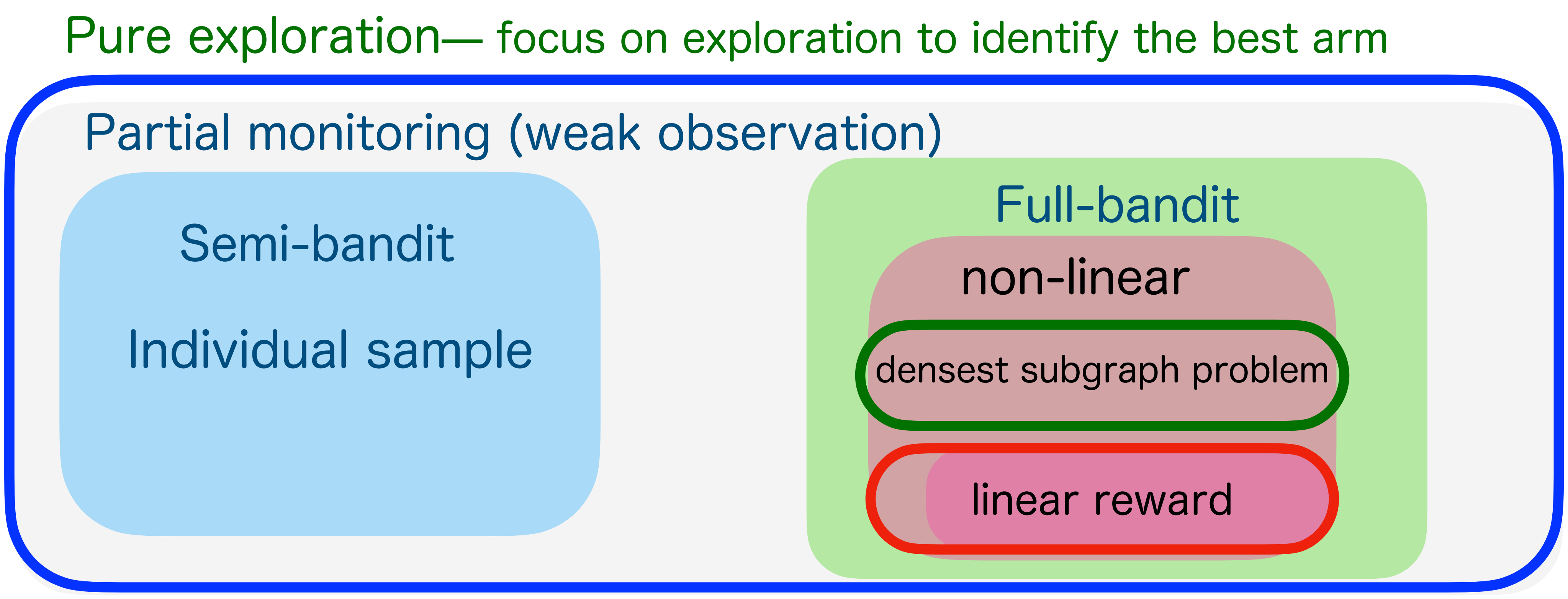
$$\mathbb{E}_\theta[\tau] \geq \log(1/2.4\delta) \min_{\lambda \in \triangle(\mathcal{X})} \max_{x \in \mathcal{X} \setminus \{x^*\}} \frac{\|x^* - x\|^2_{M(\lambda)^{-1}}}{((x^* - x)^\top \theta)^2}$$

Conclusion

To deal with uncertainty for combinatorial optimization,
we study the combinatorial bandit problems with limited feedback.

- Linear reward case with Full-bandit feedback
- Online densest subgraph discovery
- Nonlinear reward and partial-linear feedback

There are many future problems!



Thank you!