Adversarial Linear Contextual Bandits with Graph-Structured Side Observations

Lingda Wang

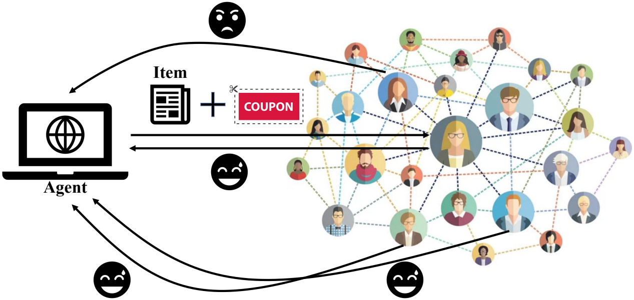
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

35th AAAI Conference on Artificial Intelligence (AAAI 2021)

February 2[nd] - 9[th], 2021

Joint work with Bingcong Li (UMN), Huozhi Zhou (UIUC), Georgios B. Giannakis (UMN), Lav R. Varshney (UIUC), and Zhizhen Zhao (UIUC)

Motivation



- Viral marketing over a social network:
  - The Agent offers a survey with a promotion to a user;
  - The user finishes the survey and share it in the social network;
  - The agent gets side observations from followers.

Linear Contextual Bandits with Side Observations

- In each time step $t = 1, 2, \dots, T$:
  - Adversary picks the feedback graph G_t, and loss vectors $\{\theta_{i,t} \in \mathbb{R}^d\}_{i=1}^K$;
  - Environment reveals the context $X_t \in \mathbb{R}^d$, $X_t \sim \mathcal{D}$;
  - The learning agent chooses action $I_t \in [K]$;
  - The agent incurs and observes loss $\ell_t(X_t, I_t) = \langle X_t, \theta_{I_t,t} \rangle \in [-1, 1]$;
  - The agent also observes losses of I_t's neighbors in G_t;
  - The adversary discloses G_t to the agent.

Linear Contextual Bandits with Side Observations

- In each time step $t = 1, 2, \ldots, T$:
  - Adversary picks the feedback graph G_t, and loss vectors $\{\theta_{i,t} \in \mathbb{R}^d\}_{i=1}^K$;
  - Environment reveals the context $X_t \in \mathbb{R}^d$, $X_t \sim \mathcal{D}$;
  - The learning agent chooses action $I_t \in [K]$;
  - The agent incurs and observes loss $\ell_t(X_t, I_t) = \langle X_t, \theta_{I_t,t} \rangle \in [-1, 1]$;
  - The agent also observes losses of I_t's neighbors in G_t;
  - The adversary discloses G_t to the agent.
- Goal: Minimize the regret:

$$\mathcal{R}_T = \max_{\pi_T \in \Pi} \mathbb{E}\left[\sum_{t=1}^{T} \left\langle X_t, \theta_{I_t,t} - \theta_{\pi_T(X_t),t} \right\rangle\right],$$

against the best policy $\pi_T \in \Pi := \{\pi_T \big| \text{all policies } \pi_T : \mathbb{R}^d \mapsto V\}$.

Existing Results on Contextual Bandits

	Regret	
	i.i.d. context	adversarial context
i.i.d. loss[1]	$\mathcal{O}(\sqrt{dKT})$	$\mathcal{O}(\sqrt{dKT})$
adversarial loss	see below	$\Theta(T)$

- For adversarial loss and i.i.d context:
  - `BISTRO++`[2]: $\mathcal{O}(T^{2/3}(K \log N)^{1/2})$, oracle-efficient;
  - `RobustLinEXP3`[3]: $\mathcal{O}(T^{2/3}(Kd \log K)^{1/3})$;
  - `RealLinEXP3`[3]: $\mathcal{O}(\sqrt{dKT \log K \log T})$.

[1] Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In Advances in Neural Information Processing Systems, pages 2312-2320.

[2] Rakhlin, A. and Sridharan, K. (2016). BISTRO: An efficient relaxation-based method for contextual bandits. In International Conference on Machine Learning, pages 1977-1985.

[3] Neu, G. and Olkhovskaya, J. (2020). Efficient and robust algorithms for adversarial linear contextual bandits. In Conference on Learning Theory, pages 1-20.

Existing Results on Contextual Bandits

	Regret	
	i.i.d. context	adversarial context
i.i.d. loss[1]	$\mathcal{O}(\sqrt{dKT})$	$\mathcal{O}(\sqrt{dKT})$
adversarial loss	see below	$\Theta(T)$

- **Question**: Is it possible to do better with side observations?

[1] Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In Advances in Neural Information Processing Systems, pages 2312-2320.

Existing Results on Contextual Bandits

	Regret	
	i.i.d. context	adversarial context
i.i.d. loss[1]	$\mathcal{O}(\sqrt{dKT})$	$\mathcal{O}(\sqrt{dKT})$
adversarial loss	see below	$\Theta(T)$

- **Question**: Is it possible to do better with side observations?
- **Answer: Yes!**

[1] Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In Advances in Neural Information Processing Systems, pages 2312-2320.

Contributions

- A new bandit model: jointly leverages contexts and side observations.
- Introducing two new algorithms, `EXP3-LGC-U` and `EXP3-LGC-IX`, with:
  - Novel loss vector estimates design;
  - Better dependence on K ($\#$ of actions).
- Promising numerical performance.

Assumptions

- **I.i.d contexts**: The context $X_t \in \mathbb{R}^d$ is drawn from a distribution $\mathcal{D}$ i.i.d., where $\mathcal{D}$ is known by the agent in advance .

Assumptions

- **I.i.d contexts**: The context $X_t \in \mathbb{R}^d$ is drawn from a distribution $\mathcal{D}$ i.i.d., where $\mathcal{D}$ is known by the agent in advance .
- **Extra observation oracle**: At each time step t, there exists an oracle that draws a context $\tilde{X}_t \in \mathbb{R}^d$ from $\mathcal{D}$, and discloses $\tilde{X}_t$ together with the losses $\tilde{l}_t(\tilde{X}_t, i)$ to the agent.

Assumptions

- **I.i.d contexts**: The context $X_t \in \mathbb{R}^d$ is drawn from a distribution $\mathcal{D}$ i.i.d., where $\mathcal{D}$ is known by the agent in advance .

- **Extra observation oracle**: At each time step t, there exists an oracle that draws a context $\tilde{X}_t \in \mathbb{R}^d$ from $\mathcal{D}$, and discloses $\tilde{X}_t$ together with the losses $\tilde{l}_t(\tilde{X}_t, i)$ to the agent.

- **Nonoblivious adversary**: The adversary can be **nonoblivious**, who is allowed to choose G_t and $\theta_{i,t}, \forall i \in V$ at time t according to arbitrary functions of the interaction history $\mathcal{F}_{t-1}$ before time step t.

Algorithm: `EXP3-LGC-U`

- Main steps in each time step t:

Algorithm: EXP3-LGC-U

- Main steps in each time step t:
  - Compute the exponential weight:

$$w_t(X_t, i) = \exp\left(-\eta \sum_{s=1}^{t-1} \left\langle X_t, \hat{\theta}_{i,s} \right\rangle\right);$$

Algorithm: EXP3-LGC-U

- Main steps in each time step t:
  - Compute the exponential weight:

$$w_t(X_t, i) = \exp\left(-\eta \sum_{s=1}^{t-1} \left\langle X_t, \hat{\theta}_{i,s} \right\rangle\right);$$

  - Draw the action $I_t \in [K]$ with probability:

$$\pi_t^a(i \middle| X_t) = (1-\gamma)\frac{w_t(X_t, i)}{\sum_{j \in V} w_t(X_t, j)} + \frac{\gamma}{K};$$

Algorithm: EXP3-LGC-U

- Main steps in each time step t:
  - Compute the exponential weight:

$$w_t(X_t, i) = \exp\left(-\eta \sum_{s=1}^{t-1} \left\langle X_t, \hat{\theta}_{i,s} \right\rangle\right);$$

  - Draw the action $I_t \in [K]$ with probability:

$$\pi_t^a(i|X_t) = (1-\gamma)\frac{w_t(X_t, i)}{\sum_{j \in V} w_t(X_t, j)} + \frac{\gamma}{K};$$

  - Estimate the loss vector $\theta_{i,t}$:

$$\hat{\theta}_{i,t} = \frac{\mathbb{I}\{i \in S_{I_t,t}\}}{q_t(i|X_t)} \Sigma^{-1} \tilde{X}_t \tilde{\ell}_t(\tilde{X}_t, i),$$

where $q_t(i|X_t) = \pi_t^a(i|X_t) + \sum_{j:j \xrightarrow{t} i} \pi_t^a(j|X_t)$.

Properties of $\hat{\theta}_{i,t}$

- Advantages and properties of $\hat{\theta}_{i,t}$:
  - The estimate is unbiased w.r.t. $\mathcal{F}_{t-1}$ and X_t:

$$\mathbb{E}\left[\hat{\theta}_{i,t} \,|X_t, \mathcal{F}_{t-1}\right] = \theta_{i,t};$$

Properties of $\hat{\theta}_{i,t}$

- Advantages and properties of $\hat{\theta}_{i,t}$:
  - The estimate is unbiased w.r.t. $\mathcal{F}_{t-1}$ and X_t:

$$\mathbb{E}\left[\hat{\theta}_{i,t} \,|X_t, \mathcal{F}_{t-1}\right] = \theta_{i,t};$$

  - Regret can be evaluated using $\hat{\theta}_{i,t}$ directly:

$$\mathcal{R}_T = \max_{\pi_T \in \Pi} \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i \in V} \left(\pi_t^a(i|X_t) - \pi_T(i|X_t)\right) \left\langle X_t, \hat{\theta}_{i,t} \right\rangle\right].$$

Main Results I

Theorem

For any positive $\eta \in (0,1)$, choosing $\gamma = \eta K\sigma^2/\lambda_{min}$, the expected cumulative regret of `EXP3-LGC-U` *satisfies:*

$$\mathcal{R}_t \leq \frac{\log K}{\eta} + \frac{2\eta K\sigma^2}{\lambda_{min}} T + \eta d \sum_{t=1}^{T} \mathbb{E}\left[Q_t\right],$$

where $Q_t = \alpha(G_t)$ if G_t is undirected, and $Q_t = 4\alpha(G_t)\log(4K^2/(\alpha(G_t)\gamma))$ if G_t is directed.

Corollary

The regret of EXP3-LGC-U is $\mathcal{O}(\sqrt{(K+\alpha(G)d)T\log K})$ in the undirected graph settting, and $\mathcal{O}(\sqrt{(K+\alpha(G)d)T}\log(KdT))$ in the directed graph setting, where $\alpha(G) \leq K$ is the averaged independence number of $\{G_t\}$.

Proof Sketch of EXP3-LGC-U

- Analyze $\log \frac{W_{t+1}(X_t)}{W_t(X_t)}$, where $W_t(x) = \sum_{i \in V} w_t(x, i)$:

$$\sum_{t=1}^{T} \mathbb{E}\left[\sum_{i \in V} \frac{\pi_t^a(i|X_t)}{1-\gamma}\left(-\eta \left\langle X_t, \hat{\theta}_{i,t} \right\rangle + \eta^2 \left\langle X_t, \hat{\theta}_{i,t} \right\rangle^2\right) + \frac{\eta\gamma}{K(1-\gamma)} \sum_{i \in V} \left\langle X_t, \hat{\theta}_{i,t} \right\rangle\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} \log \frac{W_{t+1}(X_t)}{W_t(X_t)}\right] \leq \mathbb{E}\left[-\eta \sum_{t=1}^{T} \sum_{i \in V} \pi_T(i|X_t) \left\langle X_t, \hat{\theta}_{i,t} \right\rangle - \log K\right].$$

Proof Sketch of EXP3-LGC-U

- Analyze $\log \frac{W_{t+1}(X_t)}{W_t(X_t)}$, where $W_t(x) = \sum_{i \in V} w_t(x, i)$:

$$\sum_{t=1}^{T} \mathbb{E}\left[\sum_{i \in V} \frac{\pi_t^a(i|X_t)}{1-\gamma}\left(-\eta \left\langle X_t, \hat{\theta}_{i,t} \right\rangle + \eta^2 \left\langle X_t, \hat{\theta}_{i,t} \right\rangle^2\right) + \frac{\eta\gamma}{K(1-\gamma)} \sum_{i \in V} \left\langle X_t, \hat{\theta}_{i,t} \right\rangle\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} \log \frac{W_{t+1}(X_t)}{W_t(X_t)}\right] \leq \mathbb{E}\left[-\eta \sum_{t=1}^{T} \sum_{i \in V} \pi_T(i|X_t) \left\langle X_t, \hat{\theta}_{i,t} \right\rangle - \log K\right].$$

- Reorder the terms on both sides and apply some straightforward facts:

$$\mathcal{R}_T \leq \frac{\log K}{\eta} + 2\gamma T + \eta \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i \in V} \pi_t^a(i|X_t) \left\langle X_t, \hat{\theta}_{i,t} \right\rangle^2\right]}_{A}.$$

Proof Sketch of EXP3-LGC-U

- Analyze $\log \frac{W_{t+1}(X_t)}{W_t(X_t)}$, where $W_t(x) = \sum_{i \in V} w_t(x, i)$:

$$\sum_{t=1}^{T} \mathbb{E}\left[\sum_{i \in V} \frac{\pi_t^a(i|X_t)}{1-\gamma}\left(-\eta\left\langle X_t, \hat{\theta}_{i,t}\right\rangle + \eta^2\left\langle X_t, \hat{\theta}_{i,t}\right\rangle^2\right) + \frac{\eta\gamma}{K(1-\gamma)}\sum_{i \in V}\left\langle X_t, \hat{\theta}_{i,t}\right\rangle\right]$$
$$\leq \mathbb{E}\left[\sum_{t=1}^{T} \log \frac{W_{t+1}(X_t)}{W_t(X_t)}\right] \leq \mathbb{E}\left[-\eta \sum_{t=1}^{T}\sum_{i \in V} \pi_T(i|X_t)\left\langle X_t, \hat{\theta}_{i,t}\right\rangle - \log K\right].$$

- Reorder the terms on both sides and apply some straightforward facts:

$$\mathcal{R}_T \leq \frac{\log K}{\eta} + 2\gamma T + \eta \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i \in V} \pi_t^a(i|X_t)\left\langle X_t, \hat{\theta}_{i,t}\right\rangle^2\right]}_{A}.$$

- Use graph-theoretic results form *Alon et al.* and properties of $\hat{\theta}_{i,t}$ to bound term A.

Algorithm: `EXP3-LGC-IX`

- Additional assumption: The support of $\theta_{i,t}$ and X_t is non-negative, and elements of X_t are independent.

Algorithm: EXP3-LGC-IX

- Additional assumption: The support of $\theta_{i,t}$ and X_t is non-negative, and elements of X_t are independent.
- Main steps in each time step t:

Algorithm: `EXP3-LGC-IX`

- Additional assumption: The support of $\theta_{i,t}$ and X_t is non-negative, and elements of X_t are independent.

- Main steps in each time step t:

  - Draw the action $I_t \in [K]$ proportional to the exponential weight:

$$\pi_t^a(i \big| X_t) \propto w_t(X_t, i) = \frac{1}{K} \exp\left(-\eta_t \sum_{s=1}^{t-1} \left\langle X_t, \hat{\theta}_{i,s} \right\rangle\right);$$

Algorithm: EXP3-LGC-IX

- Additional assumption: The support of $\theta_{i,t}$ and X_t is non-negative, and elements of X_t are independent.

- Main steps in each time step t:

  - Draw the action $I_t \in [K]$ proportional to the exponential weight:

$$\pi_t^a(i \big| X_t) \propto w_t(X_t, i) = \frac{1}{K} \exp\left(-\eta_t \sum_{s=1}^{t-1} \left\langle X_t, \hat{\theta}_{i,s} \right\rangle\right);$$

  - Estimate the loss vector $\theta_{i,t}$:

$$\hat{\theta}_{i,t} = \frac{\mathbb{I}\{i \in S_{I_t,t}\}}{q_t(i \big| X_t) + \beta_t} \Sigma^{-1} \tilde{X}_t \tilde{\ell}_t(\tilde{X}_t, i).$$

Property of $\hat{\theta}_{i,t}$

Claim

The loss vector estimate $\hat{\theta}_{i,t}$ in `EXP3-LGC-IX` *is optimistic:*

$$\mathbb{E}_t \left[\sum_{i \in V} \pi_t^a(i|X_t) \left\langle X_t, \hat{\theta}_{i,t} \right\rangle \middle| X_t \right]$$

$$= \sum_{i \in V} \pi_t^a(i|X_t) \left\langle X_t, \theta_{i,t} \right\rangle - \beta_t \sum_{i \in V} \frac{\pi_t^a(i|X_t)}{q_t(i|X_t) + \beta_t} \left\langle X_t, \theta_{i,t} \right\rangle,$$

where

$$0 \leq \sum_{i \in V} \frac{\pi_t^a(i|X_t)}{q_t(i|X_t) + \beta_t} \left\langle X_t, \theta_{i,t} \right\rangle \leq \sum_{i \in V} \frac{\pi_t^a(i|X_t)}{q_t(i|X_t) + \beta_t}.$$

- Control the variance of loss vector estimates.
- Regret can be evaluated using $\hat{\theta}_{i,t}$ directly.

Main Results II

Theorem

Setting $\beta_t = \sqrt{\log K / (K + \sum_{s=1}^{t-1} Q_s)}$ *and* $\eta_t = \sqrt{\log K / (dK + d\sum_{s=1}^{t-1} Q_s)}$, *the expected regret of* `EXP3-LGC-IX` *satisfies:*

$$\mathcal{R}_T \leq 2(1+\sqrt{d})\mathbb{E}\left[\sqrt{\left(K + \sum_{t=1}^{T} Q_t\right) \log K}\right],$$

for both directed and undirected graph settings, where $Q_t = 2\alpha(G_t)\log\left(1 + \frac{\lceil K^2/\beta_t \rceil + K}{\alpha(G_t)}\right) + 2$.

Corollary

The regret of EXP3-LGC-IX is $\mathcal{O}(\sqrt{\alpha(G)dT\log K\log(KT)})$ *for both directed and undirected graph settings.*

Proof Sketch of EXP3-LGC-IX

- Analyze $\log \frac{W'_{t+1}(X_t)}{W_t(X_t)}$, where $W_t(x) = \sum_{i \in V} w_t(x, i)$ and $W'_t(x) = \sum_{i \in V} \frac{1}{K} \exp(-\eta_{t-1} \sum_{s=1}^{t-1} \langle x, \hat{\theta}_{i,s} \rangle)$:

$$- \mathbb{E}\left[\frac{\log K}{\eta_{T+1}}\right] - \mathbb{E}\left[\sum_{t=1}^{T} \left\langle X_t, \hat{\theta}_{\pi_T(X_t),t} \right\rangle\right] \le \mathbb{E}\left[\sum_{t=1}^{T} \left(\frac{\log W_{t+1}(X_t)}{\eta_{t+1}} - \frac{\log W_t(X_t)}{\eta_t}\right)\right]$$

$$\le \mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{\eta_t} \log \frac{W'_{t+1}(X_t)}{W_t(X_t)}\right] \le \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i \in V} \pi_t^a(i|X_t) \left(-\left\langle X_t, \hat{\theta}_{i,t} \right\rangle + \frac{1}{2}\eta_t \left\langle X_t, \hat{\theta}_{i,t} \right\rangle^2\right)\right]}_{B}.$$

Proof Sketch of EXP3-LGC-IX

- Analyze $\log \frac{W'_{t+1}(X_t)}{W_t(X_t)}$, where $W_t(x) = \sum_{i \in V} w_t(x, i)$ and $W'_t(x) = \sum_{i \in V} \frac{1}{K} \exp(-\eta_{t-1} \sum_{s=1}^{t-1} \langle x, \hat{\theta}_{i,s} \rangle)$:

$$-\mathbb{E}\left[\frac{\log K}{\eta_{T+1}}\right] - \mathbb{E}\left[\sum_{t=1}^{T} \left\langle X_t, \hat{\theta}_{\pi_T(X_t),t} \right\rangle\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \left(\frac{\log W_{t+1}(X_t)}{\eta_{t+1}} - \frac{\log W_t(X_t)}{\eta_t}\right)\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{\eta_t} \log \frac{W'_{t+1}(X_t)}{W_t(X_t)}\right] \leq \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i \in V} \pi_t^a(i|X_t) \left(-\left\langle X_t, \hat{\theta}_{i,t} \right\rangle + \frac{1}{2}\eta_t \left\langle X_t, \hat{\theta}_{i,t} \right\rangle^2\right)\right]}_{B}.$$

- Bound term B using properties of $\hat{\theta}_{i,t}$ and graph-theoretic result from *Kocák et al.*.

Proof Sketch of EXP3-LGC-IX

- Analyze $\log \frac{W'_{t+1}(X_t)}{W_t(X_t)}$, where $W_t(x) = \sum_{i \in V} w_t(x, i)$ and $W'_t(x) = \sum_{i \in V} \frac{1}{K} \exp(-\eta_{t-1} \sum_{s=1}^{t-1} \langle x, \hat{\theta}_{i,s} \rangle)$:

$$-\mathbb{E}\left[\frac{\log K}{\eta_{T+1}}\right] - \mathbb{E}\left[\sum_{t=1}^{T} \left\langle X_t, \hat{\theta}_{\pi_T(X_t),t} \right\rangle\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \left(\frac{\log W_{t+1}(X_t)}{\eta_{t+1}} - \frac{\log W_t(X_t)}{\eta_t}\right)\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{\eta_t} \log \frac{W'_{t+1}(X_t)}{W_t(X_t)}\right] \leq \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i \in V} \pi_t^a(i|X_t) \left(-\left\langle X_t, \hat{\theta}_{i,t} \right\rangle + \frac{1}{2}\eta_t \left\langle X_t, \hat{\theta}_{i,t} \right\rangle^2\right)\right]}_{B}.$$

- Bound term B using properties of $\hat{\theta}_{i,t}$ and graph-theoretic result from *Kocák et al.*.
- Reorder the terms on both sides:

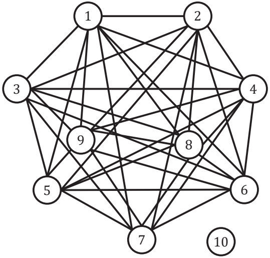
$$\mathcal{R}_T \leq 2(1 + \sqrt{d})\mathbb{E}\left[\sqrt{\left(K + \sum_{t=1}^{T} Q_t\right) \log K}\right].$$

Summary

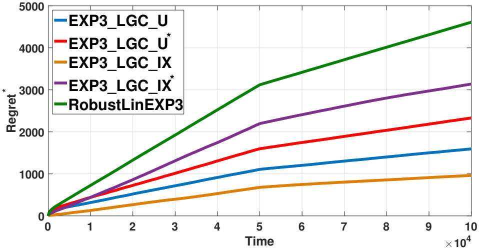
	Regret	
	Unsdirected Setting	**Directed Setting**
EXP3-LGC-U	$\mathcal{O}(\sqrt{(K+\alpha(G)d)T\log K})$	$\mathcal{O}(\sqrt{(K+\alpha(G)d)T}\log(KdT))$
EXP3-LGC-IX	$\mathcal{O}(\sqrt{\alpha(G)dT\log K\log(KT)})$	
RobustLinEXP3[2]	$\mathcal{O}(T^{2/3}(Kd\log K)^{1/3})$	
RealLinEXP3[2]	$\mathcal{O}(\sqrt{dKT\log K\log T})$	

[2] Neu, G. and Olkhovskaya, J. (2020). Efficient and robust algorithms for adversarial linear contextual bandits. In Conference on Learning Theory, pages 1 - 20.

Numerical Results



(a) Graph structure.



(b) Regret comparison.

- Experiment settings:
  - $K = 10$, $d = 10$, and $T = 10^5$;
  - $X_t(j) \sim \text{Bern}(0.5)/sqrt(d)$;
  - $\theta_{i,t}(j) = 0.1i|\cos t|/(\sqrt{d}\lceil t/50000\rceil)$.

Thank You!

- Full version of *Adversarial Linear Contextual Bandits with Graph-Structured Side Observations* is available online at: `https://arxiv.org/abs/2012.05756`.