Online Experimental Design With Estimation-Regret Trade-off Under Network Interference

Zhiheng Zhang, Zichen Wang

MAB With Interference

Network Interference: Treatment of one unit affects others → SUTVA violated.

Bandit vs Causal Inference: Bandit literature prioritizes regret; causal inference focuses on estimation accuracy.
Offline designs ensure power but incur regret in sequential settings.

Our Goal: Unify both! Design policies with a provable estimation-regret Pareto frontier.

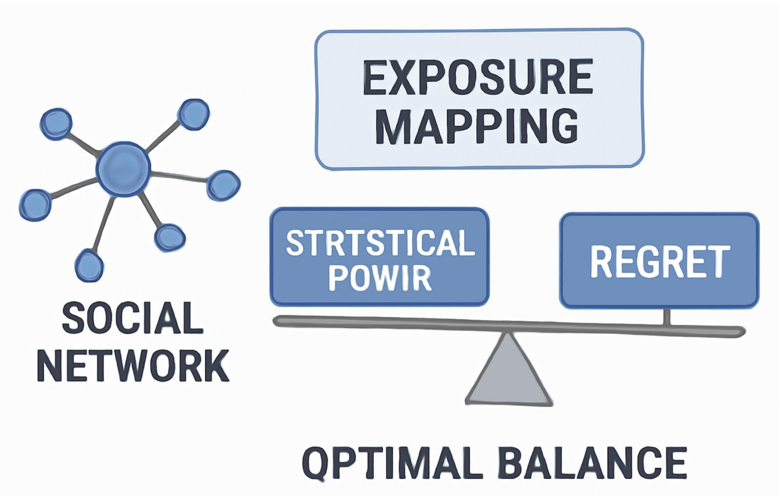
Basic Setting:

There exists a node set U with N nodes and each node has a K-armed set $\mathbf{K}$.

Exposure Mapping:

$$s = \mathbf{S}(i, A, H),$$

where $\mathbf{S}: U \times \mathbf{K}^U \times \{0,1\}^{N \times N} \to U_s$, $A = (a_1, \cdots, a_N) \in \mathbf{K}^U$, H denotes the adjacency matrix, and $|U_s| = d_s$. s denotes the exposure arm.



Learning Goals

Iteration processes of MAB with interference:

for t = 1:T
  Select a exposure super arm $S_t \in U_\mathcal{E}$
  Sample $A_t \in \mathbf{K}^U$ based on S_t (see Algorithm 2)
  for node i∈[N]
    Pull arm $a_{i,t} \in \mathbf{K}$
  end
  Receive reward feedback $\{Y_{i,t}(A_t)\}_{i=1}^N$
end

Theoretical Contributions

Pareto frontier

Regret based on exposure mapping. According to the action space reduction in Eq (3), we provide a more general and realistic regret compared to Jia et al. (2024); Simchi-Levi and Wang (2024); Agarwal et al. (2024) (refer to Example 1-4). We define the clustering set $\mathcal{C} := \{\mathcal{C}_q\}_{q\in[C]}, C = |\mathcal{C}|$ where $\forall i \neq j, i, j \in [C], \mathcal{C}_i \cap \mathcal{C}_j = \varnothing, \cup\{\mathcal{C}_q\}_{q\in[C]} = \mathcal{U}$. For brevity, we denote $\mathcal{C}^{-1}(i)$ as the cluster of node i. We define the exposure-based regret:

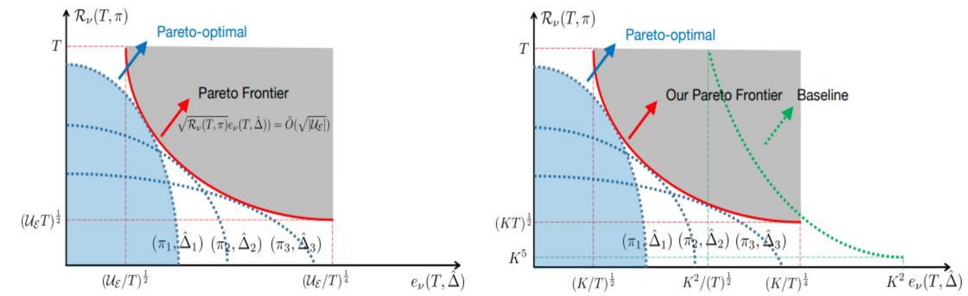
$$\mathcal{R}_\nu(T,\pi) = \frac{T}{N}\sum_{i\in\mathcal{U}} \tilde{Y}_i(S^*) - \frac{1}{N}\mathbb{E}_\pi\left[\sum_{t\in[T]}\sum_{i\in\mathcal{U}} \tilde{r}_{i,t}(S_t)\right], \quad S^* = \arg\max_{S\in\mathcal{U}_\mathcal{E}} \sum_{i\in\mathcal{U}} \tilde{Y}_i(S), \tag{4}$$

$$\min_{\{\pi,\hat{\Delta}\}} \max_{\nu\in\mathcal{E}_0} \left(\mathcal{R}_\nu(T,\pi), e_\nu(T,\hat{\Delta})\right), \text{ where } e_\nu(T,\hat{\Delta}) := \max_{S_i,S_j\in\mathcal{U}_\mathcal{E}} \mathbb{E}\left[\left|\Delta^{(i,j)} - \hat{\Delta}_T^{(i,j)}\right|\right].$$

Theorem 1 *Given any* $\mathbf{S}$ *and* $\mathcal{C}$ *that satisfies Condition 1. Given any online decision-making policy* π*, the trade-off between the regret and the estimation exhibits*

$$\inf_{\hat{\Delta}_T} \max_{\nu\in\mathcal{E}_0} \left(\sqrt{\mathcal{R}_\nu(T,\pi)}\, e_\nu(T,\hat{\Delta})\right) = \Omega_{K,T}\left(\sqrt{|\mathcal{U}_\mathcal{E}|}\right). \tag{6}$$

Visualization



(a) Our general result under interference. (b) The comparison with the baseline without interference.

Results

Unified Online Design with Network Interference

We introduce MAB-N, a unified framework for online design under network interference, utilizing exposure mapping to bridge the multi-objective minimax trade-off.

Achieving Pareto-Optimality: Our framework achieves Pareto-optimality between treatment effect estimation and regret efficiency under network interference. We propose specific criteria for MAB algorithms to reach this optimal trade-off.

UCB-TSN Algorithm: We propose the UCB-TSN algorithm to achieve the Pareto trade-off, with validated performance through experiments. It outperforms existing methods in (i) the degenerated single-unit case and (ii) the extended adversarial bandit setting.