

MOTIVATION AND CONTRIBUTION

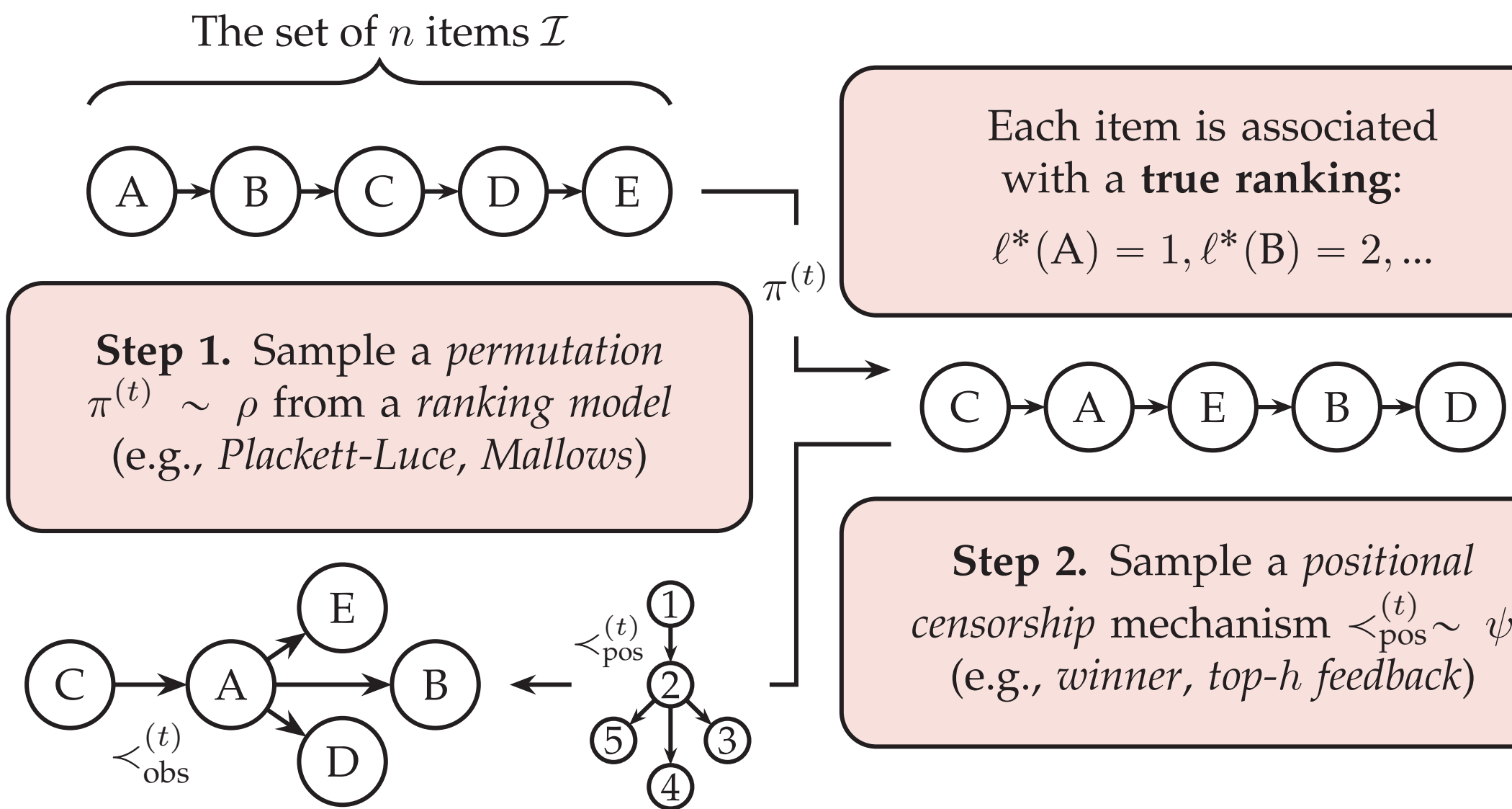
Incomplete rankings are ubiquitous in real-world datasets, but *theoretically grounded* algorithms typically rely on **rigid modeling assumptions**:

- On the **latent ranking model** (e.g., Plackett-Luce (PL), Mallows)
- On the **observed ranking portion** (e.g., winner, top- h feedback)

⇒ We introduce a **unified framework** for incomplete rankings, showing that **learning** is possible **beyond classical assumptions**

SETTING

DATA



ASSUMPTION

Complete Consensus. [2] For any $1 \leq l < m \leq n$, if $\pi(l) < \pi(m)$ then:

$$\rho(\pi) > \rho(\pi \circ \tau_{l,m}),$$

where $\pi \circ \tau_{l,m}$ is equal to π , except for l and m that are *swapped*

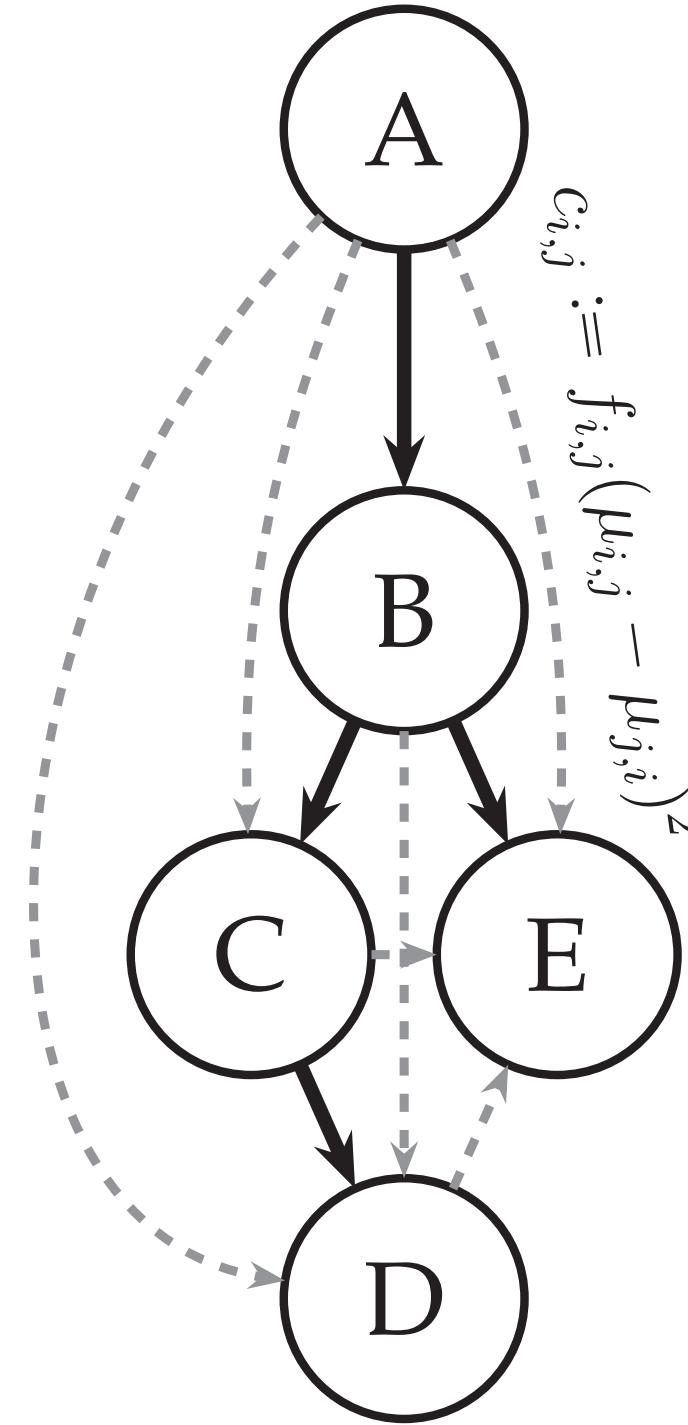
GOAL AND PERFORMANCE INDEX

The *learning objective* is to **identify** w.p. $1 - \delta$ the **top- k** items:

$$i_1^*, \dots, i_k^* \text{ where } i_l^* := \ell^{*-1}(l)$$

We *evaluate* algorithms in terms of **sample complexity**: $S^{(\delta)}(\text{Alg})$

STATISTICAL COMPLEXITY



The *sample complexity* to learn a **preference** $i < j$ is governed by:

- The **pref. prob.** $p_{i,j} := \mathbb{P}_{\rho,\psi}[i <_{\text{obs}} j]$
- The **frequency** $f_{i,j} := p_{i,j} + p_{j,i}$
- The **cond. pref. prob.** $\mu_{i,j} := p_{i,j}/f_{i,j}$

Consider an *arborescence* $\mathcal{A}$ that connects i_1^* to $i_2^*, \dots, i_{k-1}^*$ to i_k^* and i_k^* to all the others

The **bottleneck identification rate** [1] is $\min_{(i,j) \in \mathcal{A}} c_{i,j}$. Let $c_k^* = \max_{\mathcal{A}} \min_{(i,j) \in \mathcal{A}} c_{i,j}$

Theorem 4.3. For every Alg, δ :

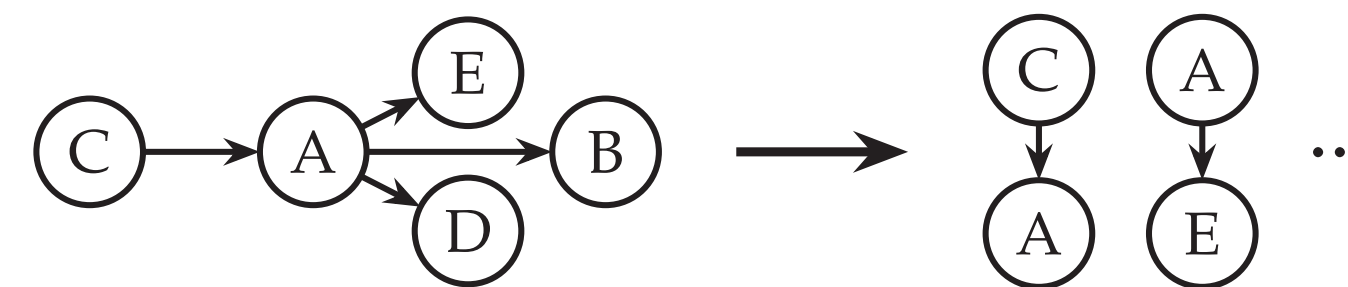
$$S^{(\delta)}(\text{Alg}) = \Omega(1/c_k^* \log(1/\delta))$$

Corollary 4.4. $c_{i,j} \approx \langle H_{\rho, i,j}, P_{\psi} \rangle_F$ with P_{ψ} being the *expected adjacency matrix* of $<_{\text{pos}}$

ALGORITHM

PATCHING INCOMPLETE RANKINGS THROUGH TRANSITIVE EDGES (PIRATE)

Rank Breaking

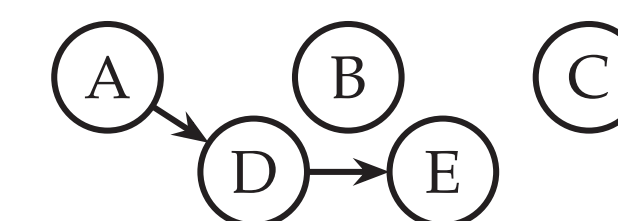


Estimator

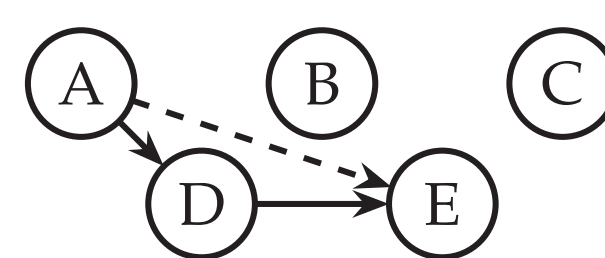
$$\begin{aligned} \# \text{ Wins } w_{i,j} &= \sum_{t=1}^T \mathbb{I}[i <_{\text{obs}}^{(t)} j] \\ \# \text{ Comp. } N_{i,j} &= w_{i,j} + w_{j,i} \\ \text{Ratio } \hat{\mu}_{i,j} &= w_{i,j}/N_{i,j} \end{aligned}$$

Dominance Graph

Add edge when $\mu_{i,j} > 1/2$ w.p. $1 - \delta$



Transitive Closure



Sample Complexity

Stop when $\mathcal{A} \subseteq \text{dominance graph}$

Theorem 4.2. For every δ :

$$S^{(\delta)}(\text{PIRATE}) = \tilde{O}(1/c_k^* \log(1/\delta))$$

NUMERICAL SIMULATIONS

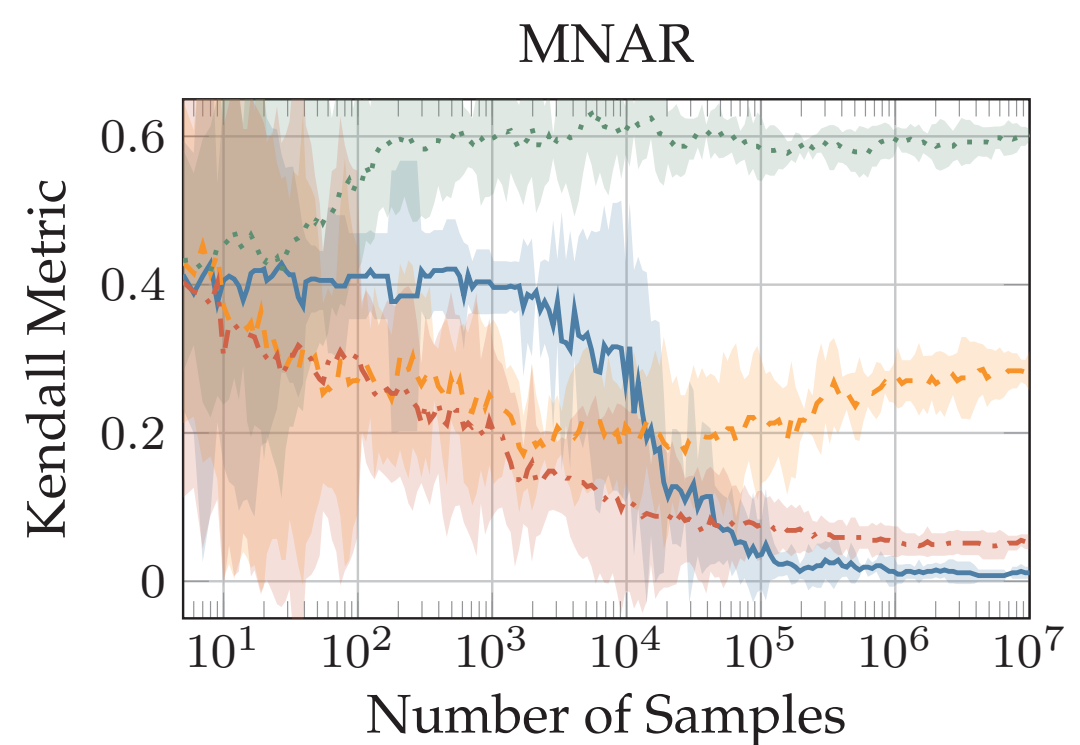
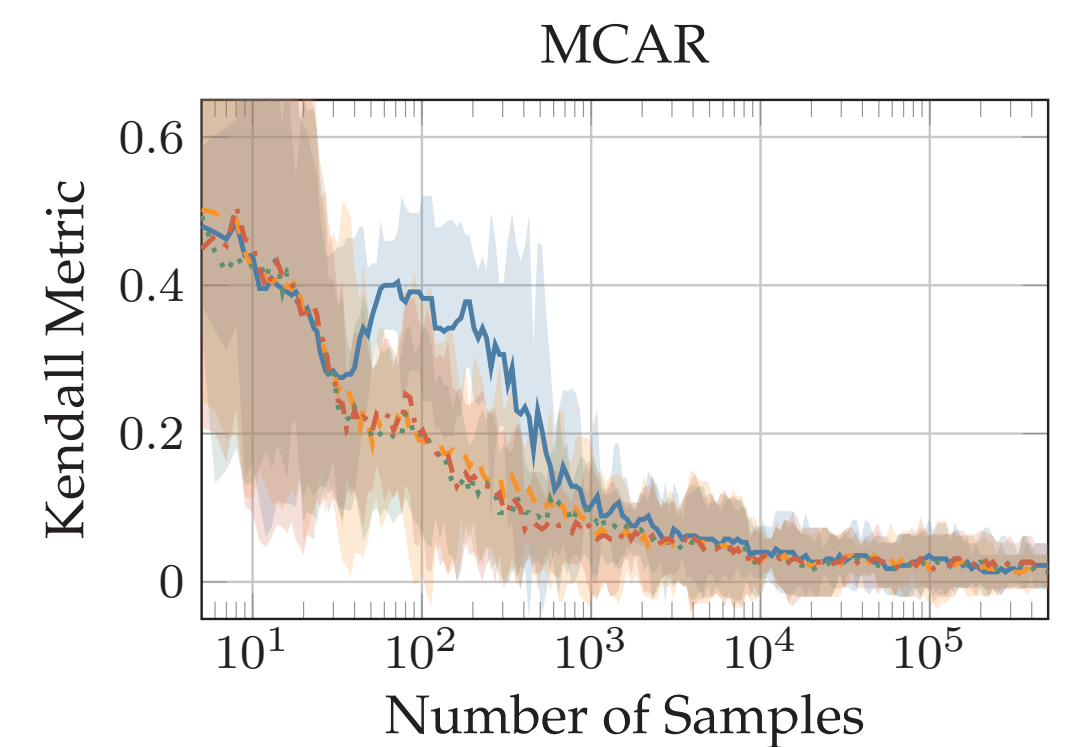
PIRATE vs. state-of-the-art **LTR** baselines:

- **RANK CENTRALITY**: Markov-chain stationary distribution
- **BORDA COUNT**: fraction of won comparisons
- **PL PAIRWISE MLE**: MLE of the PL parameters

SETTING. Latent ranking from **Plackett-Luce** ranking model

MCAR: a random pair is observed

MNAR: only the *top-3* / *bottom-3* order



Shaded: 90% CIs over 5 runs

MCAR: PIRATE **matches** the specialized baselines
MNAR: it is the **only** method recovering the true ranking

REFERENCES

- [1] Paolo M. Camerini. The Min-Max spanning tree problem and some extensions. *Information Processing Letters*, 1978.
- [2] Douglas E. Critchlow, Michael A. Fligner, and Joseph S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 1991.