

---

# Robust Learning to Rank from Incomplete Rankings under Positional Censoring

---

**Cristiano Migali**   Gianmarco Genalti   Alberto Maria Metelli   Marco Mussi

`cristiano.migali@polimi.it`

Politecnico di Milano



**POLITECNICO**  
MILANO 1863



# Motivation

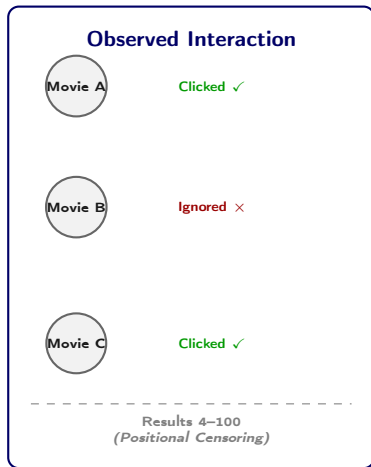
Real dataset often present **incomplete rankings**

E.g., in **Recommender Systems**:

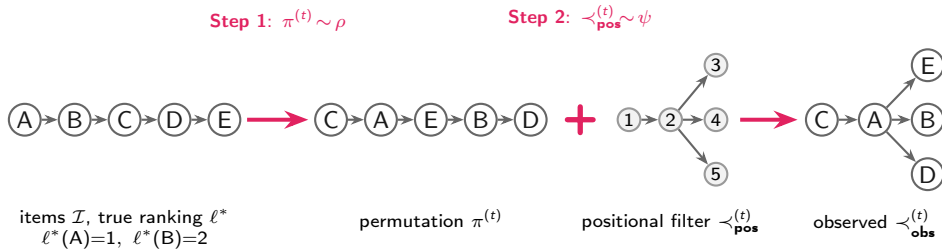
Users interact and rate only some of the items  
(the items that they very like/dislike)

However, *theoretically grounded* algorithms often rely on **rigid assumptions**:

- ▶ **Latent ranking model** (e.g., Plackett–Luce, Mallows)
- ▶ **Observed ranking portion** (e.g., top- $h$  feedback only, MCAR comparisons)



## Data generation process



### Assumption — Complete Consensus

(Critchlow et al., 1991)

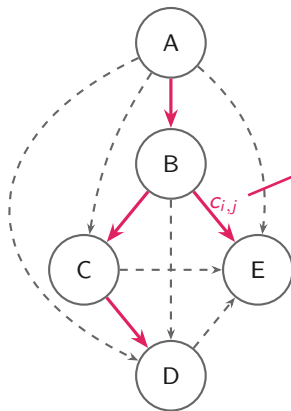
### Goal & Performance Index

Identify the **top- $k$**  items.

**Performance index:**

sample complexity  $S^{(\delta)}(\text{Alg})$

# Statistical Complexity



## Capacity

$$c_{i,j} = f_{i,j} (\mu_{i,j} - \mu_{j,i})^2$$

$f_{i,j}$ : prob. of observing a comparison of  $i$  and  $j$

$\mu_{i,j}$ : prob. that  $i$  is preferred to  $j$ , given they are compared

## Sample Complexity Lower Bound

**Thm. 4.3**  $\forall \text{Alg}, \delta$ :

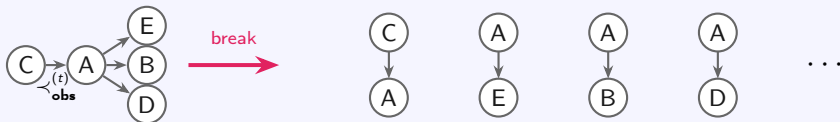
$$S^{(\delta)}(\text{Alg}) = \Omega\left(\frac{1}{c_k^*} \log \frac{1}{\delta}\right)$$

where  $c_k^*$  is the **bottleneck capacity** (Camerini, 1978) to connect the top- $k$  items in order, then the  $k$ -th to the rest

# Algorithm: Pirate

## Patching Incomplete Rankings through Transitive Edges

### Rank Breaking



### Estimator

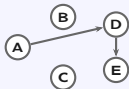
$$w_{i,j} = \sum_t \mathbb{I}[i \prec_{\text{obs}}^{(t)} j]$$

$$N_{i,j} = w_{i,j} + w_{j,i}$$

$$\hat{\mu}_{i,j} = w_{i,j} / N_{i,j}$$

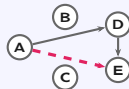
### Dominance Graph

add edge if  
 $\mu_{i,j} > \frac{1}{2}$  w.p.  $1 - \delta$



### Transitive Closure

add implied  
(dashed) edges



### Sample Complexity

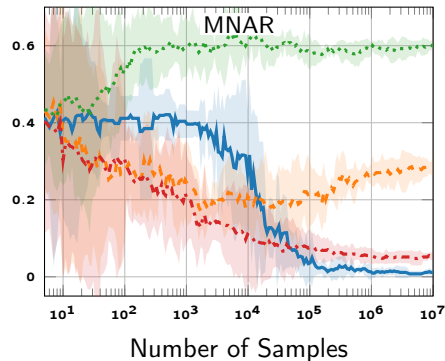
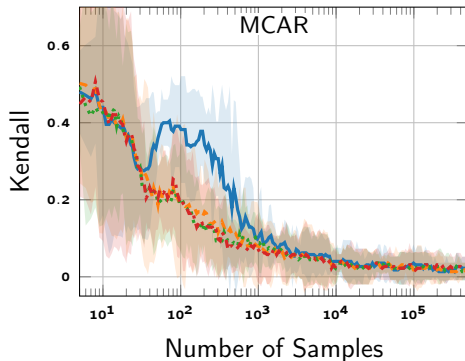
**Thm. 4.2**  $\forall \delta$ :

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

# Numerical Simulations

**Pirate** vs. **RANK CENTRALITY**, **BORDA COUNT**, **PL PAIRWISE MLE**.

Latent **Plackett–Luce**; **MCAR**: random pair | **MNAR**: top-3/bottom-3. Shaded: 90% CIs, 5 runs



**MCAR:** **PIRATE** **matches** the specialized baselines

**MNAR:** it is the **only** method recovering the true ranking

# References

---

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