Adversarial top-*K* ranking

Vincent Tan
ECE and Mathematics, NUS

2016 Shannon Workshop, SJTU
December 15, 2016

Changho Suh @KAIST

Renbo Zhao @NUS

Ranking

One of the fundamental problems in a wide range of contexts.

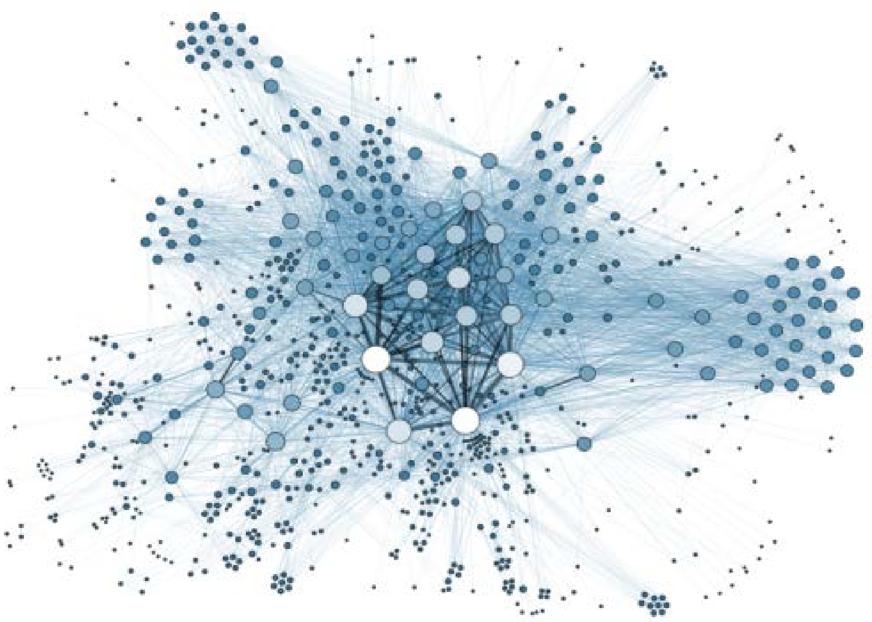
Applications: web search, recommendation systems, social choice, sports competitions, voting, etc

Lots of efforts made in developing ranking algorithms.

A variety of statistical **models** introduced for evaluating ranking algorithms.

Ranking : An Example and Difficulties

Example: web search



$n = 10^9$ websites

$\binom{n}{2} \approx n^2 = 10^{18}$ comparisons

Really need $\geq n^2$ comparisons?

Large Scale Ranking

Suppose (i) we want a total ordering &
(ii) pairwise comparisons are randomly given (probabilistically).

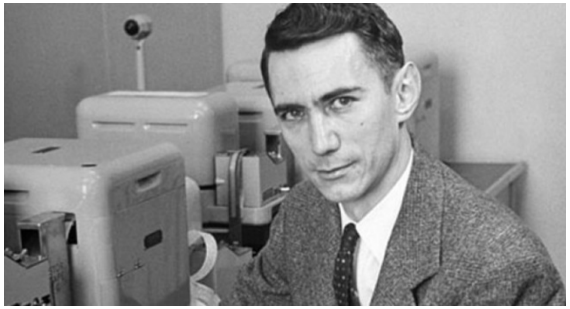
→ Requires $\geq n^2$ comparisons.

No way to identify the ordering between 1 and 2 without the direct comparison.

Comparison must be made with probability 1.

Things are even worse if one has noisy data.

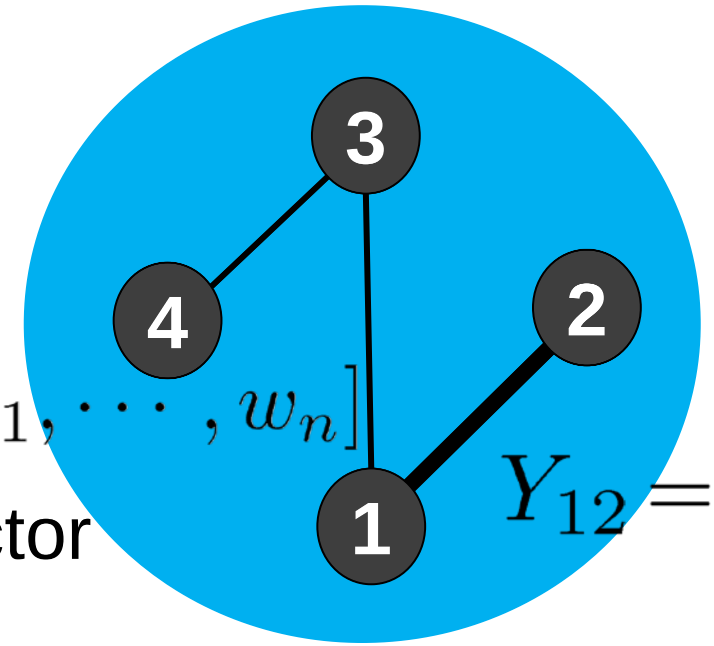
Solution: Shannon-theoretic approach

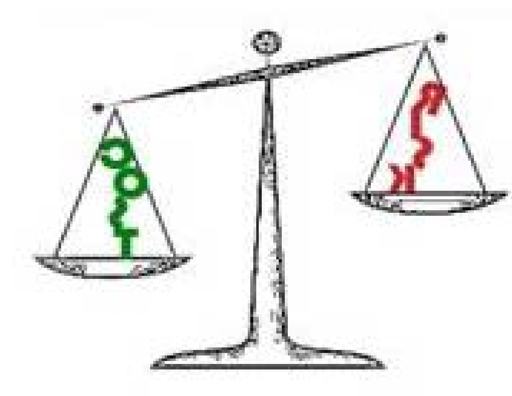


A prominent model

Bradley-Terry-Luce '52

comparison graph





$$\boldsymbol{w} = [w_1, \cdots, w_n]$$

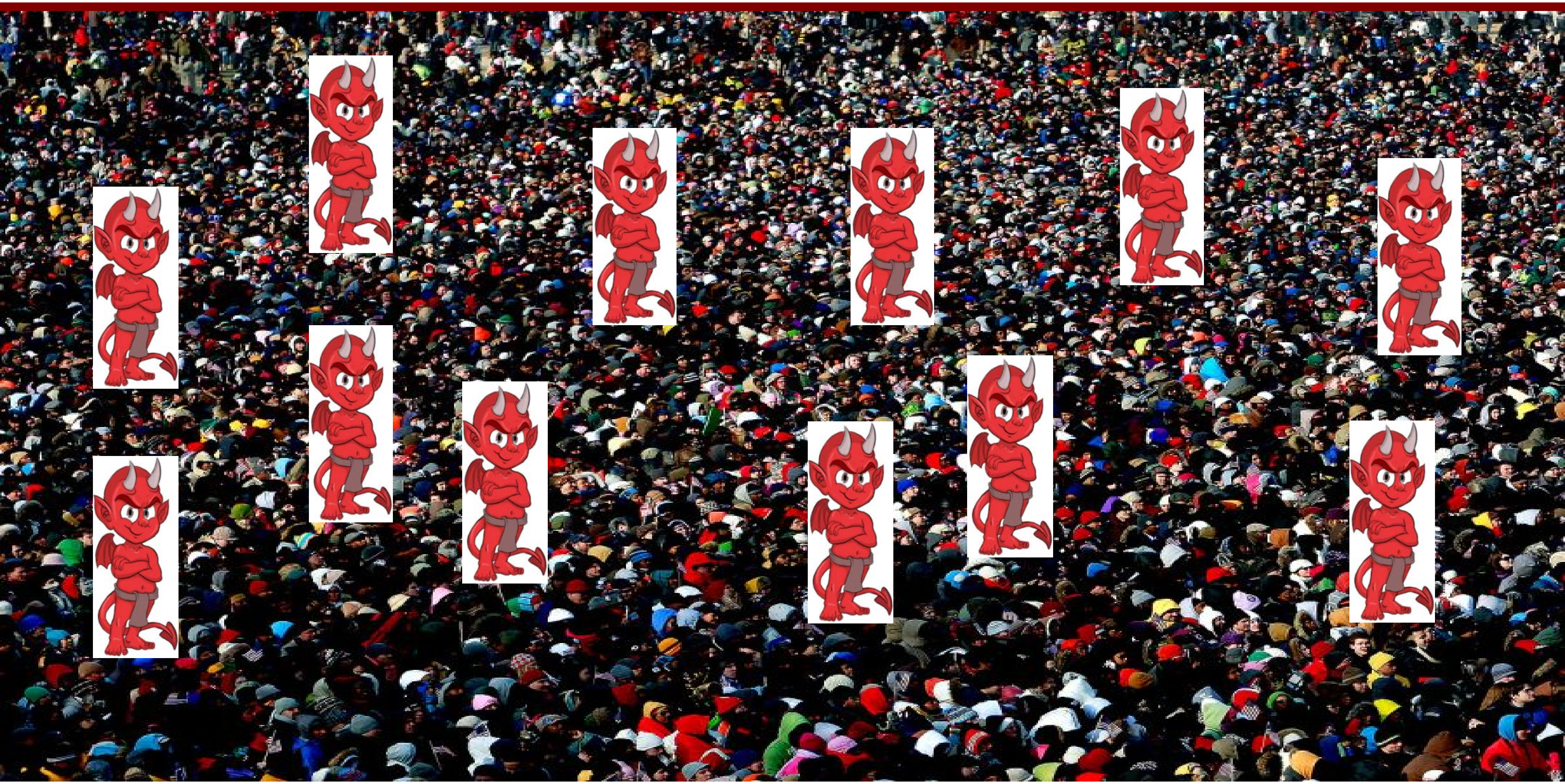
score vector

$$Y_{12} = \mathbb{I}\{\text{item } 1 \succ \text{item } 2\}$$

$$\sim \text{Bern}\left(\frac{w_1}{w_1 + w_2}\right)$$

L independent copies: $Y_{12}^{(1)}, \cdots, Y_{12}^{(L)}$

Challenge in crowdsourced settings



Spammers provide answers in an adversarial manner.

The BTL does not respect the adversarial scenario.

Adversarial BTL model

Adapted from Chen et al. '13

faithful population η portion

$$Y_{ij} \sim \text{Bern}\left(\frac{w_i}{w_i + w_j}\right) \quad \text{w.p. } \eta$$

adversarial population $1-\eta$ portion

$$Y_{ij} \sim \text{Bern}\left(\frac{w_j}{w_i + w_j}\right) \quad \text{w.p. } 1-\eta$$

$$Y_{ij} \sim \text{Bern}\left(\eta \cdot \frac{w_i}{w_i + w_j} + (1-\eta) \cdot \frac{w_j}{w_i + w_j}\right)$$

WLOG assume $\eta \in (0.5, 1]$

Related work: Crowdsouced BTL

[Chen et al. '13]

Given an observed pair, each sample has different distributions.

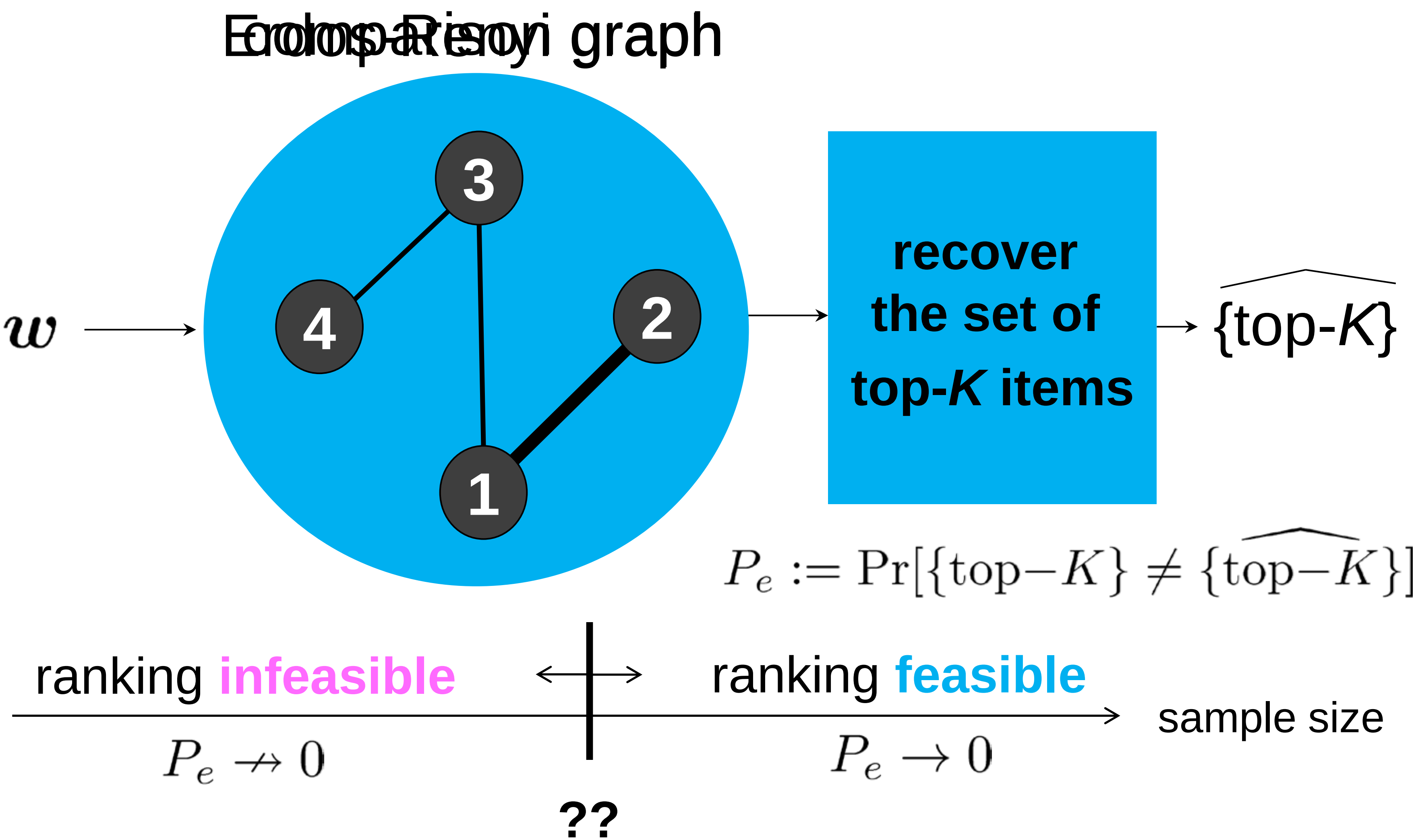
$$Y_{ij}^{(\ell)} \sim \text{Bern}\left(\eta_\ell \cdot \frac{w_i}{w_i + w_j} + (1 - \eta_\ell) \cdot \frac{w_j}{w_i + w_j} \right)$$

quality parameter

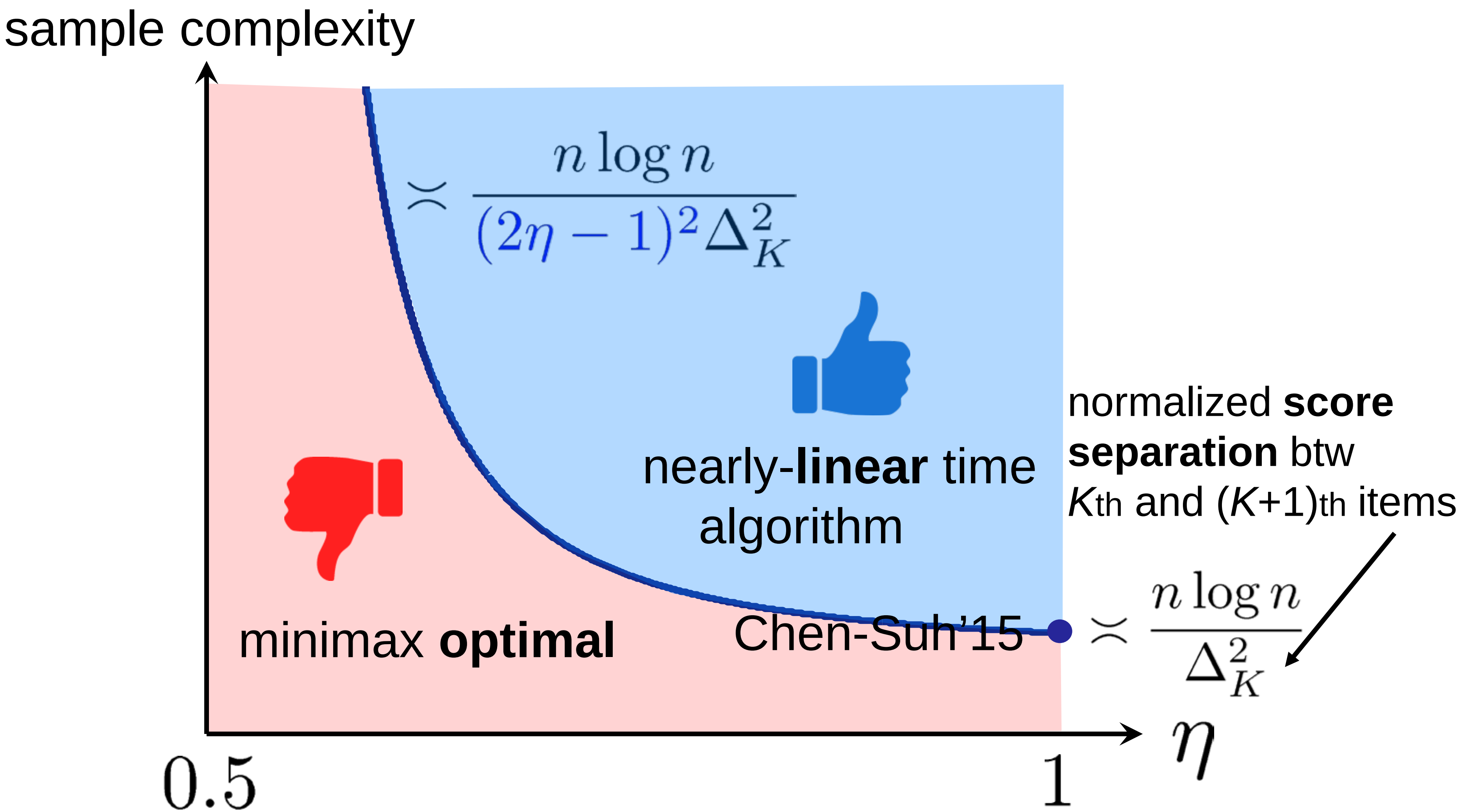
Subsumes as a special case our adversarial BTL model.

Developed a ranking algorithm **but without theoretical guarantees.**

Top-*K* ranking

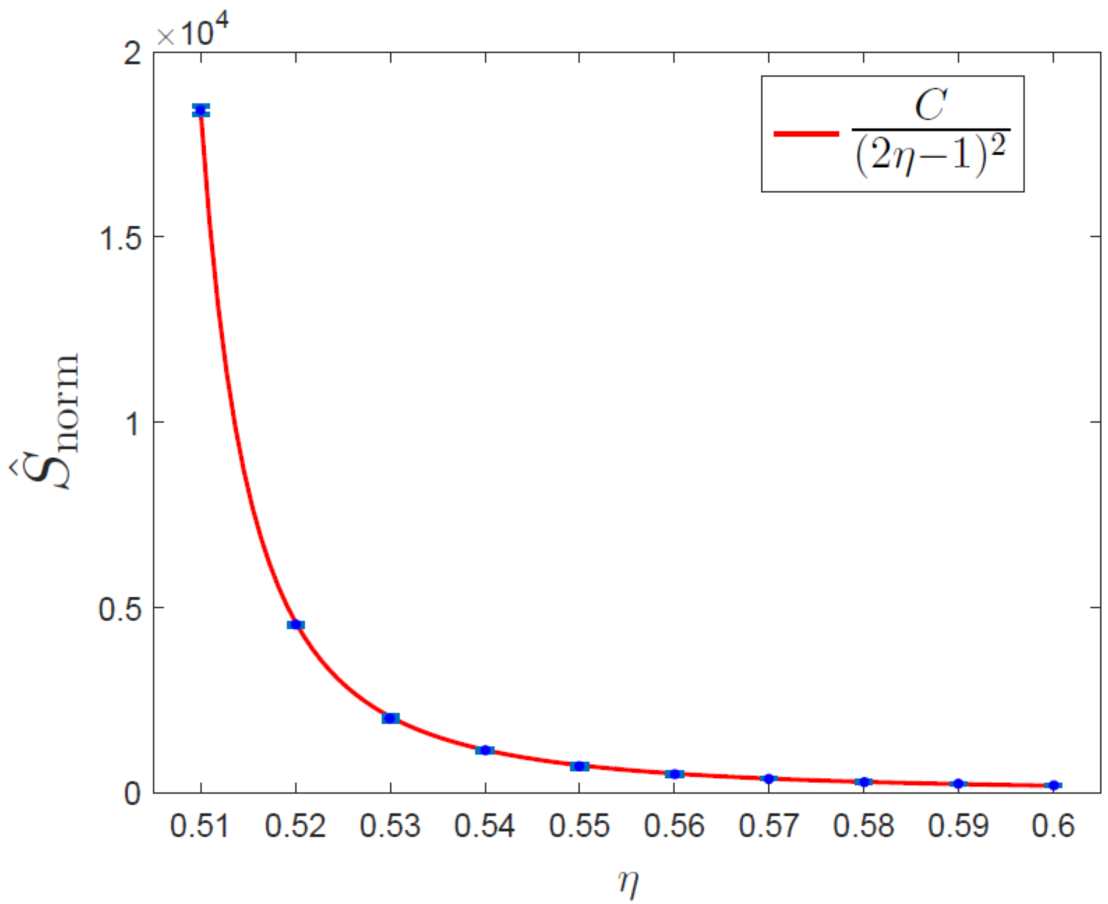


Contribution #1 η-known



Contribution #1 η-known

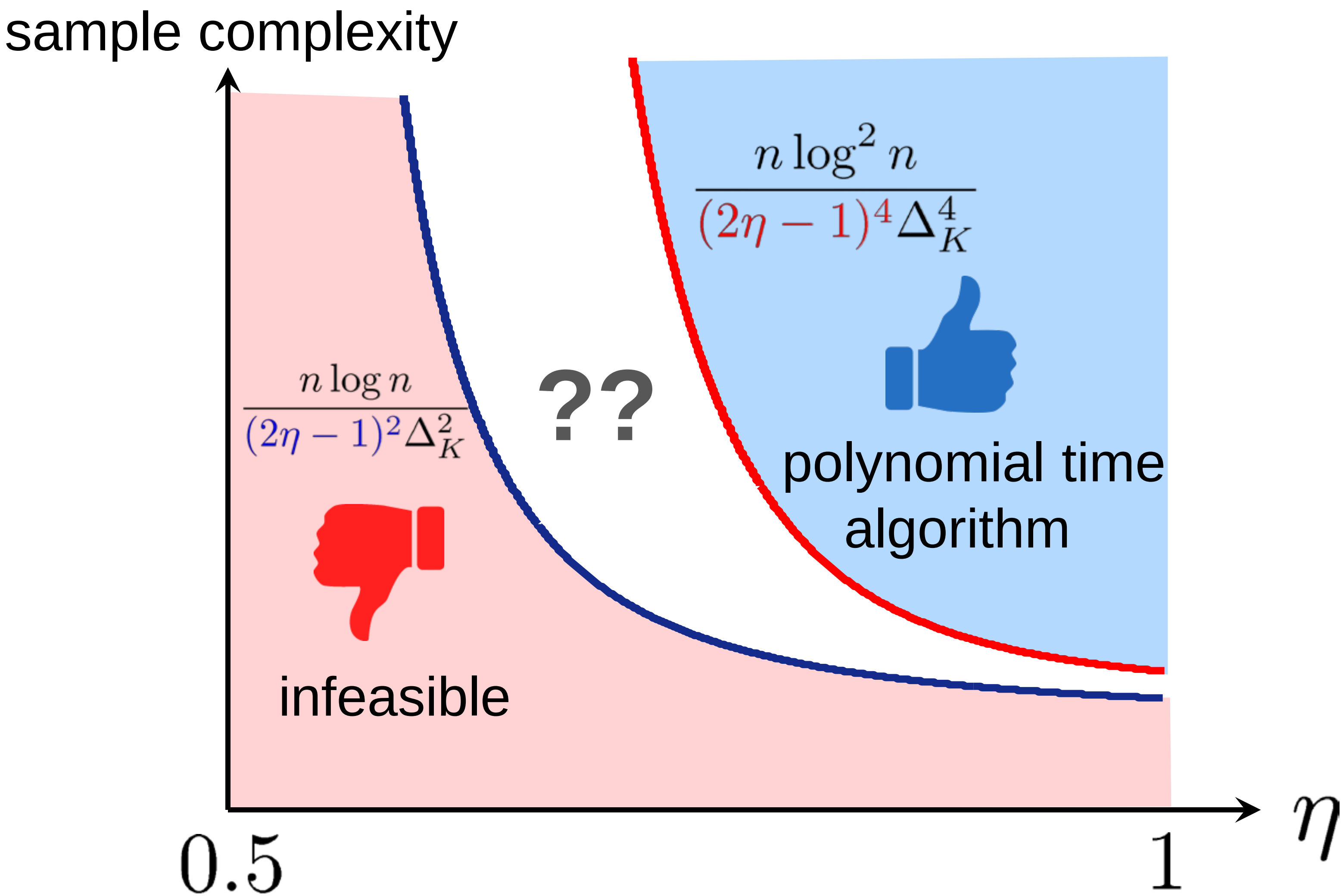
Experimental Result



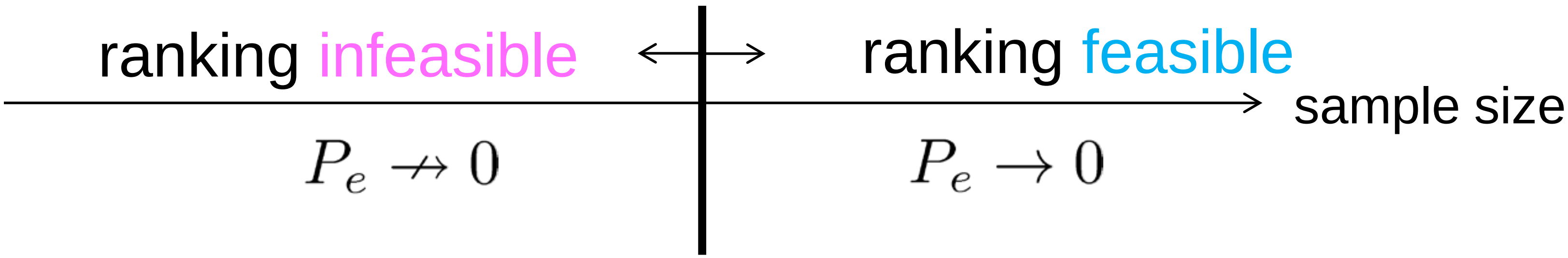
$$\hat{S} := \binom{n}{2} p \hat{L}_{\rm ave}$$

$$\hat{S}_{\rm norm} := \frac{\hat{S}}{(n \log n)/\Delta_K^2}$$

Contribution #2 η-unknown



Optimality

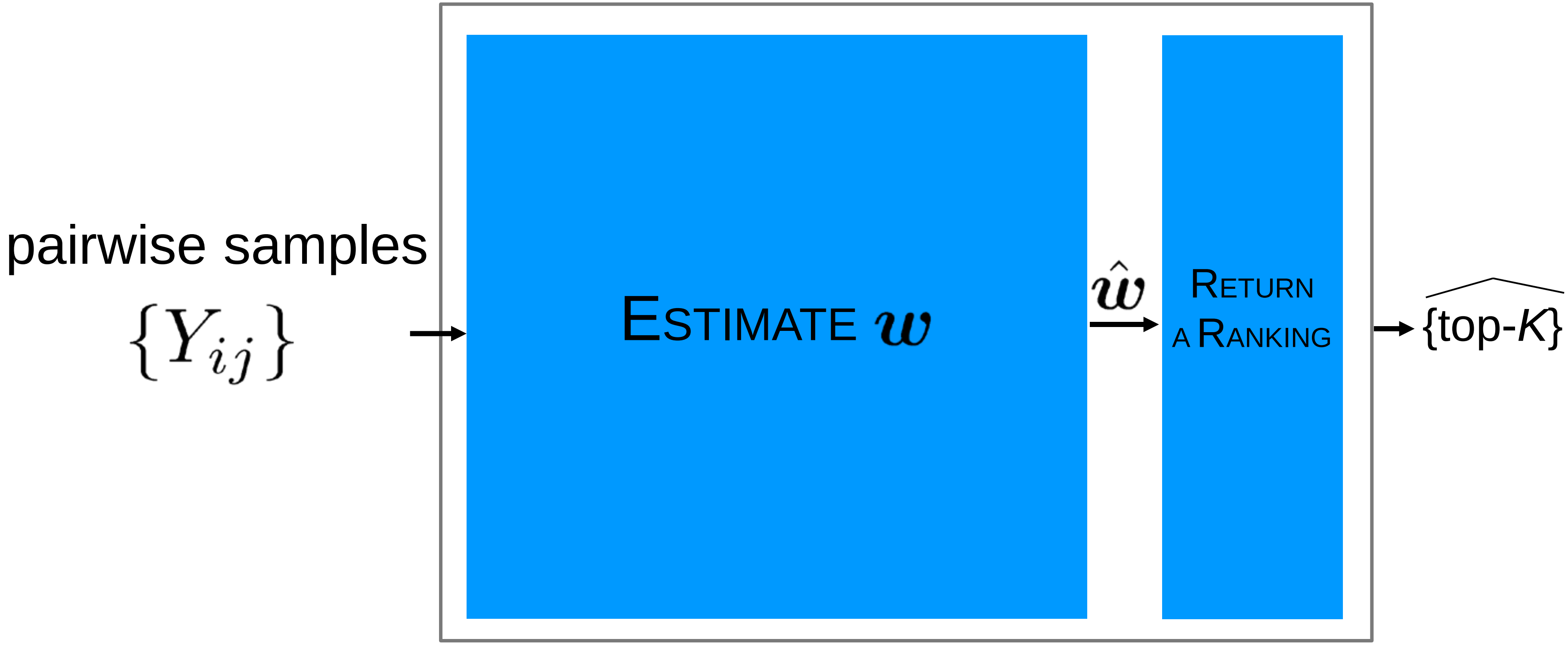


$$\asymp \frac{n \log n}{(2\eta - 1)^2 \Delta_K^2}$$

- Minimax optimality:
  Construction of the worst-case score vector
- Translation to hypothesis testing:
  Construction of multiple hypotheses

Tools: KL divergence, Generalized Fano's inequality,
Reverse Pinsker inequality

Ranking algorithm η-known



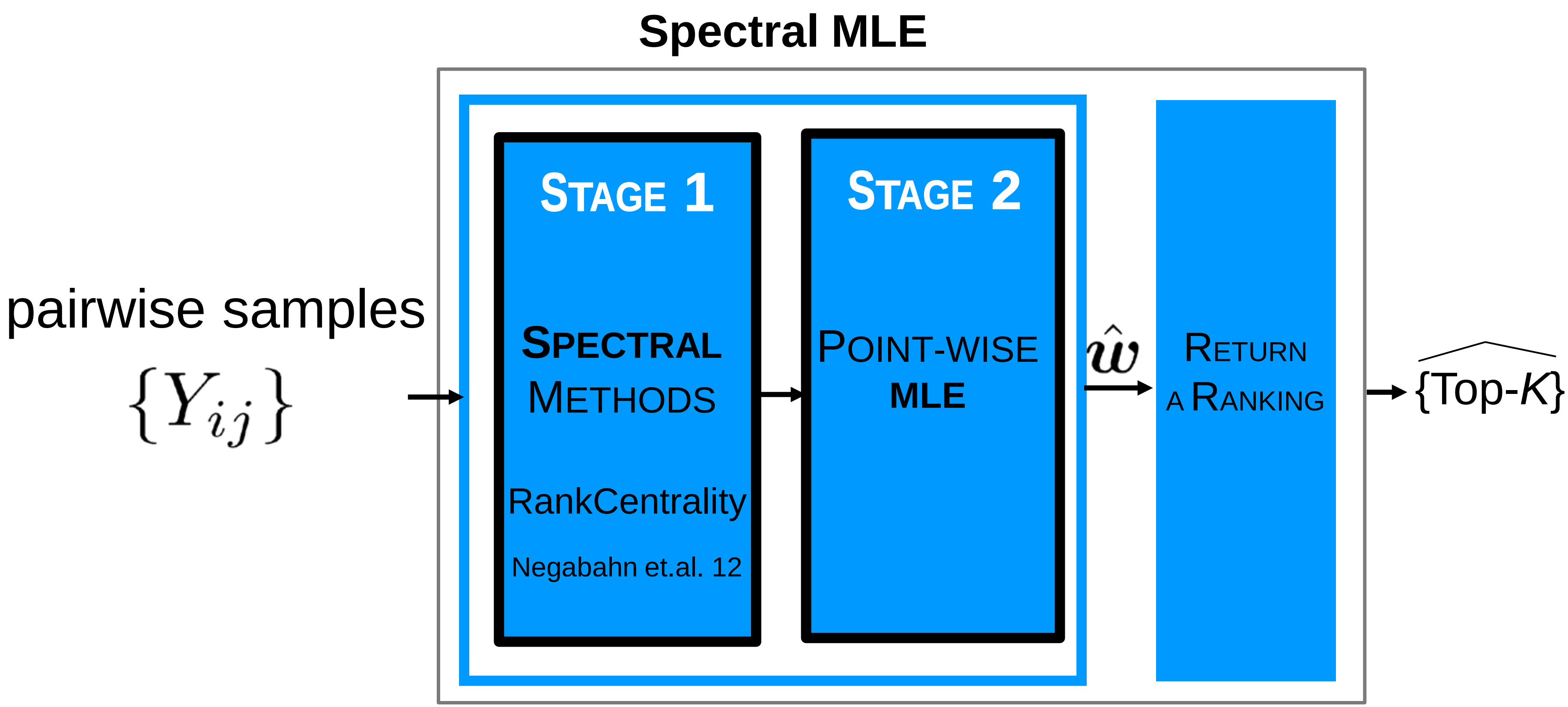
Remember: Scores determine a ranking.

→ Take a **two-step** approach.

Ranking algorithm η-known

$\eta = 1$

Chen-Suh '15



key message: small **MSE** → small **point-wise error** → high **ranking accuracy**

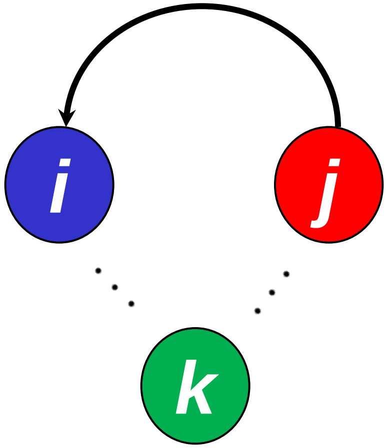
How to ensure small MSE? $\eta \in (0.5, 1]$

Recall $\eta = 1:$

L ind. copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \cdots, Y_{ij}^{(L)}$

$$\frac{1}{L}\sum_{\ell=1}^{L} Y_{ij}^{(\ell)} \longrightarrow \frac{w_i}{w_i + w_j}$$

$$\frac{1}{L}\sum_{\ell=1}^{L} Y_{ij}^{(\ell)}$$



stationary dist. → $\boldsymbol{w}$ (up to const. scaling)

How to ensure small MSE? $\eta \in (0.5, 1]$

Arbitrary η

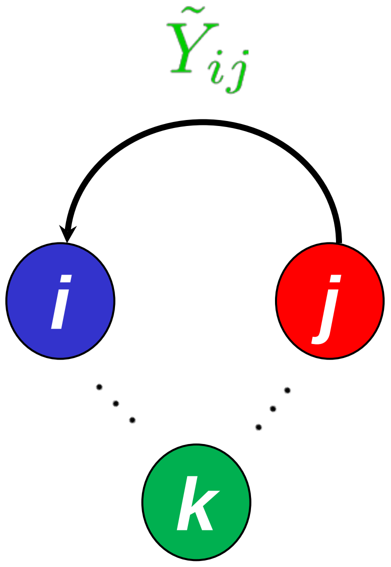
L ind. copies $Y_{ij}^{(1)}, Y_{ij}^{(2)}, \cdots, Y_{ij}^{(L)}$

$$\frac{1}{L}\sum_{\ell=1}^{L} Y_{ij}^{(\ell)} \longrightarrow \eta \cdot \frac{w_i}{w_i + w_j} + (1-\eta) \cdot \frac{w_j}{w_i + w_j}$$

$$= (2\eta - 1) \cdot \frac{w_i}{w_i + w_j} + 1 - \eta$$

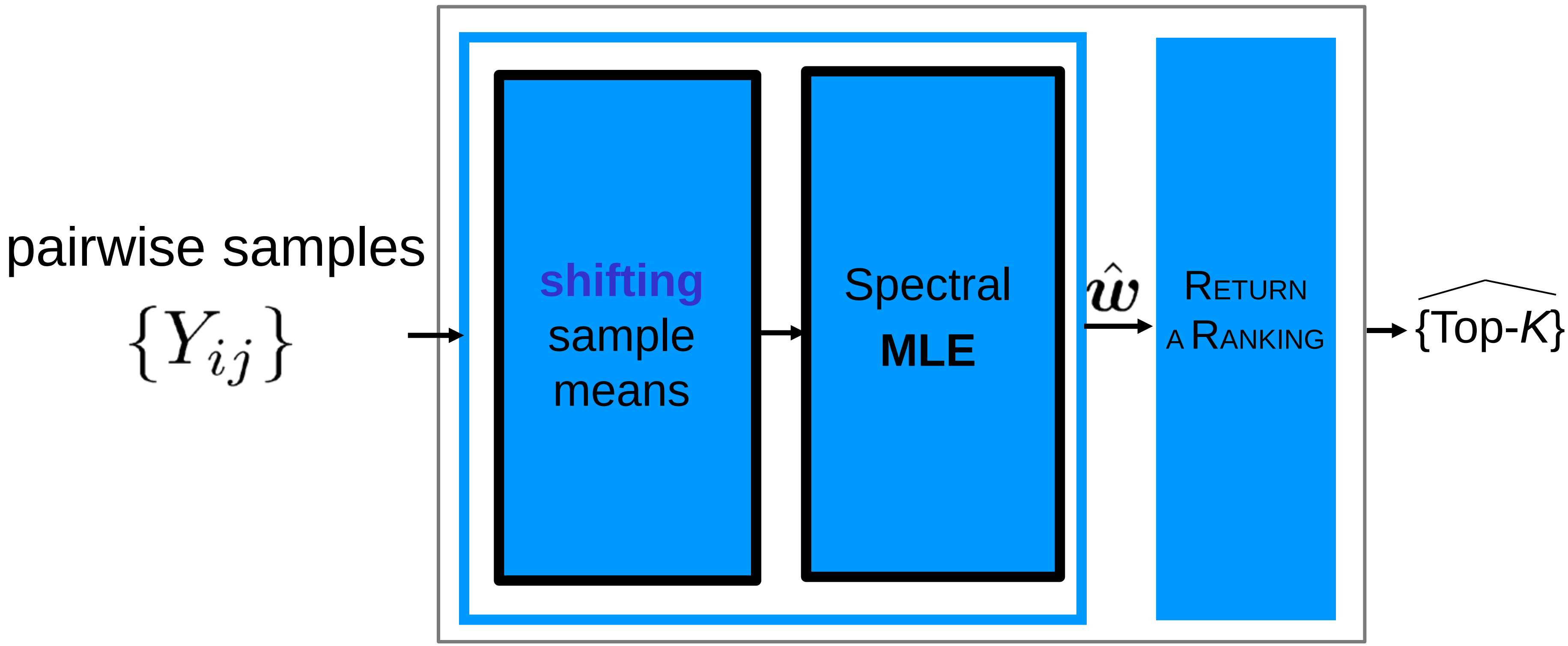
$\tilde{Y}_{ij} :=$

$$\frac{\frac{1}{L}\sum_{\ell=1}^{L} Y_{ij}^{(\ell)} - (1-\eta)}{2\eta - 1} \longrightarrow \frac{w_i}{w_i + w_j}$$



Idea: Construct Markov Chain now with $\{\tilde{Y}_{ij}\}$

Ranking algorithm η-known $\eta \in (0.5, 1]$



Using several ineq. (Hoeffding, Bernstein, Tropp …), showed:

sample size $\gtrsim \frac{n \log n}{(2\eta - 1)^2 \Delta_K^2}$ → feasible ranking

What if η is unknown?

Adversarial BTL model is a **mixture model**.

Mixture model learning problems are difficult in general.

Recent development:

Tensor methods: Jain-Oh ‘14, Anandkumar et al. ‘14

Key insight: exact 2^{nd} & 3^{rd} moments → sufficient statistics

Our setting:

Can obtain estimates of 2^{nd} & 3^{rd} moments

→ Can estimate η

Unknown η? High-level Algorithm

1) Turn weights into **distribution vectors**

$$\pi_0 := \left[\cdots \quad \frac{w_i}{w_i + w_j} \quad \frac{w_j}{w_i + w_j} \quad \frac{w_{i'}}{w_{i'} + w_{j'}} \quad \frac{w_{j'}}{w_{i'} + w_{j'}} \quad \cdots\right]^T$$

2) Estimate **moments**

$$M_2 := \eta \pi_0 \otimes \pi_0 + (1 - \eta)\pi_1 \otimes \pi_1,$$

$$M_3 := \eta \pi_0 \otimes \pi_0 \otimes \pi_0 + (1 - \eta)\pi_1 \otimes \pi_1 \otimes \pi_1,$$

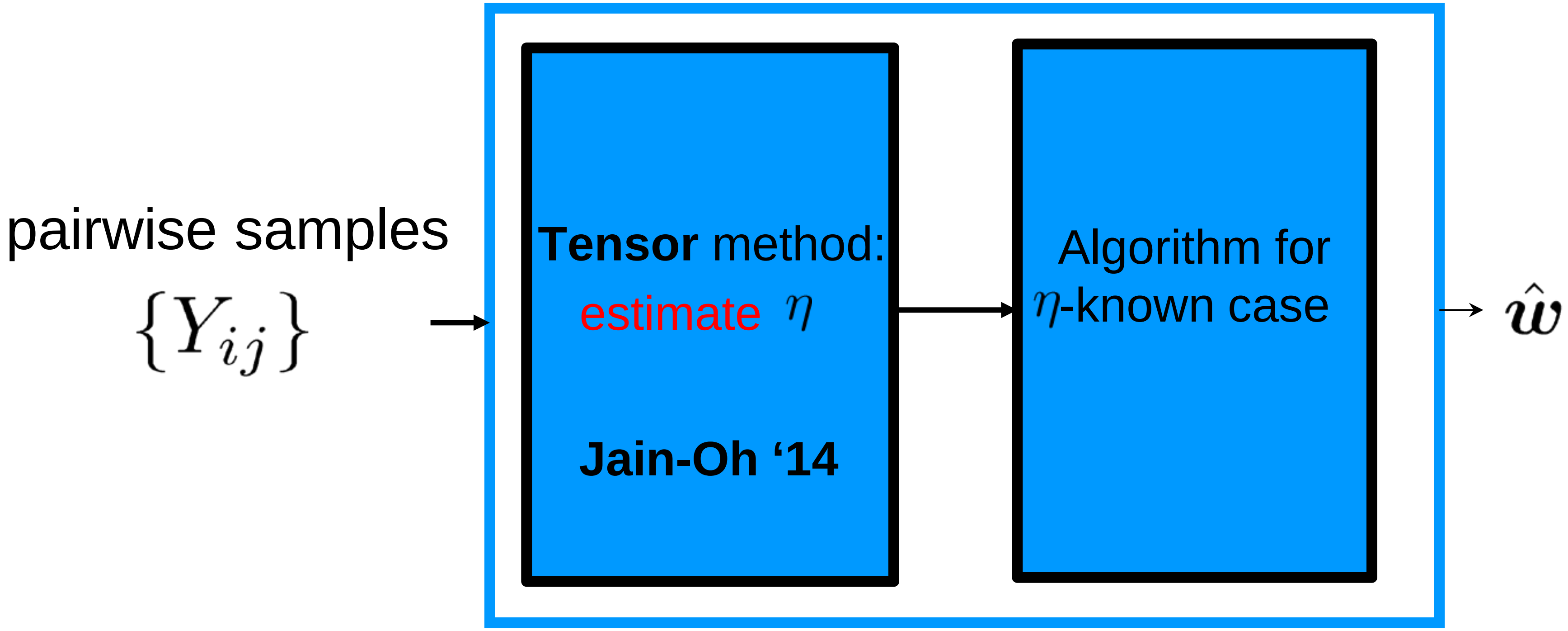
3) Solve a **Least Squares Problem**

$$\widehat{G} \in \underset{Z \in \mathbb{R}^{2\times 2\times 2}}{\arg\min} \left\| \mathcal{P}_{\Omega_3} \left(Z \left[P_{\widehat{M}_2} \right]_3 - \frac{1}{|\mathcal{I}_2|} \sum_{t \in \mathcal{I}_2} \otimes^3 \underline{Y}^{(t)} \right) \left[Q_{\widehat{M}_2} \right]_3 \right\|_F^2 .$$

4) Find **leading eigenvalue** λ_1 of $\widehat{G}$ which is related to the mixing weight as follows:

$$\hat{\eta} = \lambda_1^{-2}$$

Ranking algorithm η-unknown



How does the η-estimation affect sample complexity?

Tradeoff btw $|\eta - \hat{\eta}|$ & sample complexity

With very careful analysis, we can derive a lemma:

$$|\eta - \hat{\eta}| \leq \epsilon \quad \Longrightarrow \quad \text{sample size required} \gtrsim \frac{n \log^2 n}{\epsilon^2}$$

$|\eta - \hat{\eta}| \downarrow$ ⟹ $\|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_\infty \downarrow$ but $\text{sample size required} \uparrow$

$|\eta - \hat{\eta}| \uparrow$ ⟹ $\text{sample size required} \downarrow$ but $\|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_\infty \uparrow$

We could find a **<u>sweet spot</u>** to show that

sample size $\gtrsim \dfrac{n \log^2 n}{(2\eta - 1)^4 \Delta_K^4}$ ⟹ feasible ranking

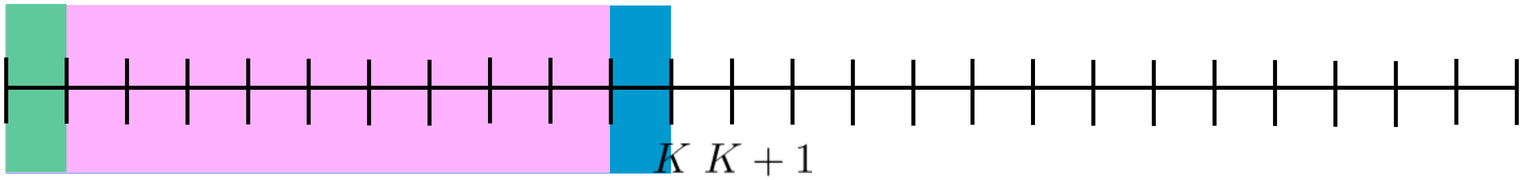
Conclusion

- Explored a **top-*K*** ranking problem for an **adversarial** setting.
- Characterized order-wise optimal **sample complexity** for η-known case.
- Established an **upper bound** on the sample complexity for η-unknown case.
- Developed computationally **efficient** algorithms.
- Full version of paper on arXiv 1602.04567.

backup slides

Extension

- Detailed ranking among the top-*K* items



$$\text{sample size required} \gtrsim \frac{n \log n}{\Delta_K^2}$$

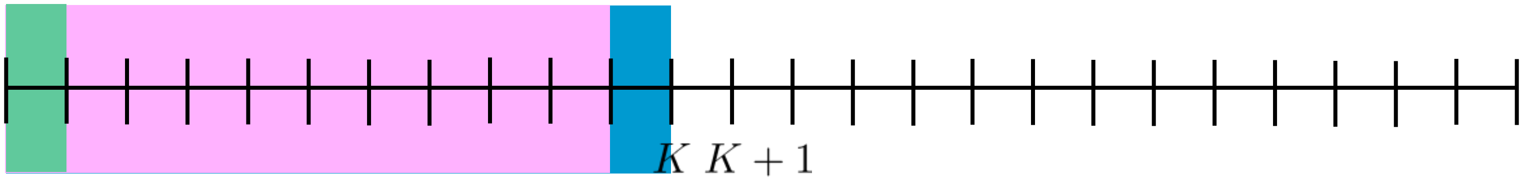
$$\text{sample size required} \gtrsim \frac{n \log n}{\Delta_{K-1}^2}$$

$$\vdots$$

$$\text{sample size required} \gtrsim \frac{n \log n}{\Delta_1^2}$$

Extension

- Detailed ranking among the top-*K* items



$$\text{sample size required} \gtrsim \frac{n \log n}{\min \Delta_i^2}$$

- Can easily extend to any-*K* ranking.

Related work: General comp. model

[Shah-Wainwright`15]

No ground truth score vector

Instead we are given: $p_{ij} := \Pr[\text{item } i \succ \text{item } j]$

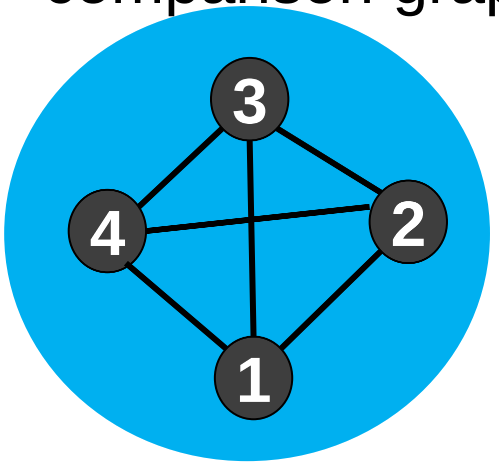
Subsumes as special cases many parametric models (including BTL and ours)

Assumption:

# of comparisons per every edge

$\sim \text{Binom}(r, p)$

comparison graph



Random vs. adaptive sampling

Case 1: $\Delta_1 \approx \Delta_2 \approx \cdots \approx \Delta_{n-1}$

Both sampling methods yield almost the same performance.

Case 2: $\Delta_1 \ll \Delta_i \; (i \geq 2)$

Adaptive sampling outperforms random sampling.