Spectral Method and Regularized MLE Are Both Optimal for Top-K Ranking

Cong Ma

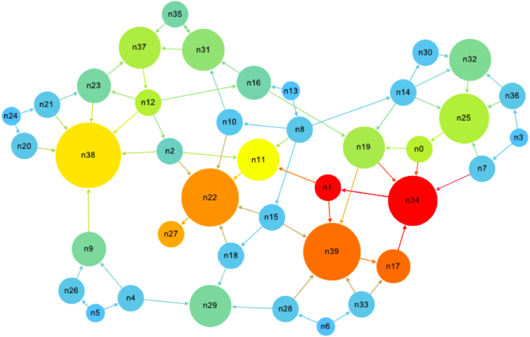
ORFE, Princeton University

Joint work with Yuxin Chen, Jianqing Fan and Kaizheng Wang

Ranking

A fundamental problem in a wide range of contexts

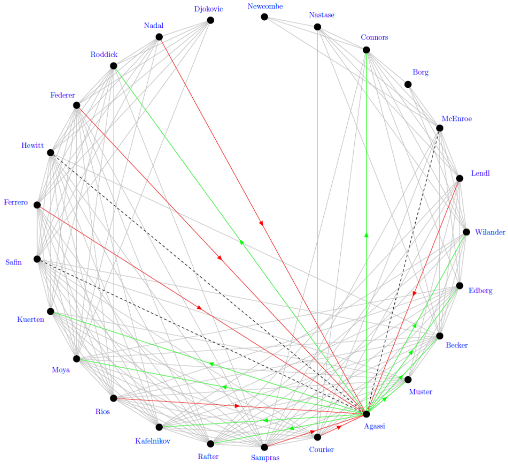
- web search, recommendation systems, admissions, sports competitions, voting, ...



PageRank

figure credit: Dzenan Hamzic

Rank aggregation from pairwise comparisons

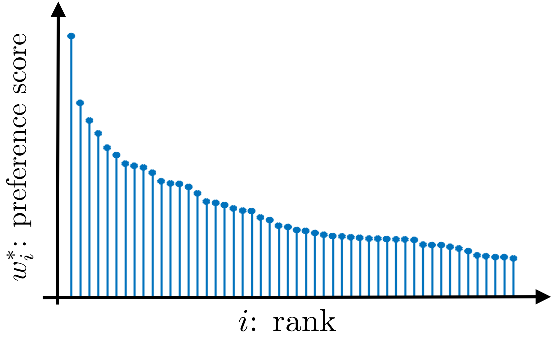


pairwise comparisons for ranking top tennis players

figure credit: Bozóki, Csató, Temesi

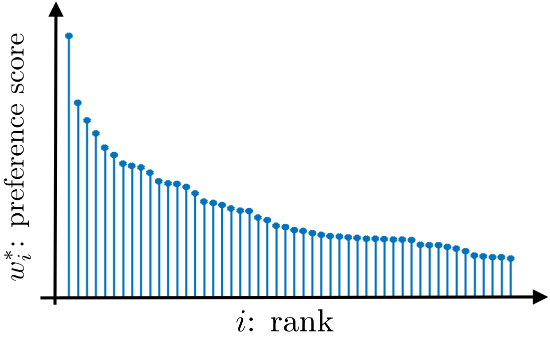
Parametric models

Assign latent preference score to each of n items $\boldsymbol{w}^* = [w_1^*, \cdots, w_n^*]$



Parametric models

Assign latent preference score to each of n items $\boldsymbol{w}^* = [w_1^*, \cdots, w_n^*]$



- **This work:** Bradley-Terry-Luce model: for $\boldsymbol{w}^* \in \mathbb{R}_+^n$

$$\mathbb{P}\{\text{item } j \text{ beats item } i\} = \frac{w_j^*}{w_i^* + w_j^*}$$

Other parametric models

- Thurstone model: for $\boldsymbol{w}^* \in \mathbb{R}^n$

$$\mathbb{P}\left\{\text{item } j \text{ beats item } i\right\} = \underbrace{\Phi}_{\text{Gaussian cdf}} \left(w_j^* - w_i^*\right)$$

Other parametric models

- Thurstone model: for $\boldsymbol{w}^* \in \mathbb{R}^n$

$$\mathbb{P}\left\{\text{item } j \text{ beats item } i\right\} = \underbrace{\Phi}_{\text{Gaussian cdf}} \left(w_j^* - w_i^*\right)$$

- Parametric models: for nondecreasing $f : \mathbb{R} \to [0,1]$ which obey

$$f(t) = 1 - f(-t), \qquad \forall t \in \mathbb{R}$$

Then we set

$$\mathbb{P}\left\{\text{item } j \text{ beats item } i\right\} = f\left(w_j^* - w_i^*\right)$$

Typical ranking procedures

Estimate latent scores

$\longrightarrow$ rank items based on score estimates

Top-K ranking

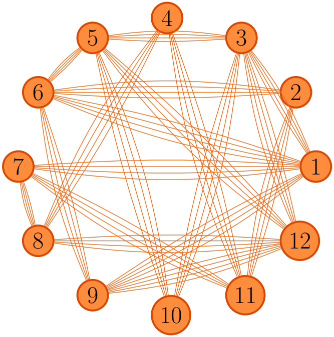
Estimate latent scores

$\longrightarrow$ rank items based on score estimates

Goal: identify the set of top-K items with pairwise comparisons

Model: random sampling

- Comparison graph: Erdős–Rényi graph $\mathcal{G} \sim \mathcal{G}(n, p)$

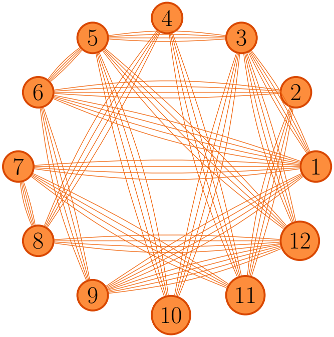


- For each $(i, j) \in \mathcal{G}$, obtain L paired comparisons

$$y_{i,j}^{(l)} \overset{\text{ind.}}{=} \begin{cases} 1, & \text{with prob. } \frac{w_j^*}{w_i^* + w_j^*} \\ 0, & \text{else} \end{cases} \qquad 1 \le l \le L$$

Model: random sampling

- Comparison graph: Erdős–Rényi graph $\mathcal{G} \sim \mathcal{G}(n, p)$



- For each $(i, j) \in \mathcal{G}$, obtain L paired comparisons

$$y_{i,j} = \frac{1}{L} \sum_{l=1}^{L} y_{i,j}^{(l)} \qquad \text{(sufficient statistic)}$$

Spectral method (Rank Centrality)

Negahban, Oh, Shah '12

- Construct a probability transition matrix $\boldsymbol{P} = [P_{i,j}]_{1 \le i,j \le n}$:

$$P_{i,j} = \begin{cases} \frac{1}{d} y_{i,j}, & \text{if } (i,j) \in \mathcal{E}, \\ 1 - \frac{1}{d} \sum_{k:(i,k)\in\mathcal{E}} y_{i,k}, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

- Return score estimate as leading left eigenvector of $\boldsymbol{P}$

Rationale behind spectral method

In large-sample case, $\boldsymbol{P} \xrightarrow{L\to\infty} \boldsymbol{P}^* = [P^*_{i,j}]_{1\le i,j\le n}$:

$$P^*_{i,j} = \begin{cases} \frac{1}{d}\frac{w_j^*}{w_i^*+w_j^*}, & \text{if } (i,j)\in\mathcal{E}, \\ 1-\frac{1}{d}\sum_{k:(i,k)\in\mathcal{E}}\frac{w_k^*}{w_i^*+w_k^*}, & \text{if } i=j, \\ 0, & \text{otherwise.} \end{cases}$$

Rationale behind spectral method

In large-sample case, $\boldsymbol{P} \xrightarrow{L \to \infty} \boldsymbol{P}^* = [P^*_{i,j}]_{1 \le i,j \le n}$:

$$P^*_{i,j} = \begin{cases} \frac{1}{d} \frac{w_j^*}{w_i^* + w_j^*}, & \text{if } (i,j) \in \mathcal{E}, \\ 1 - \frac{1}{d} \sum_{k:(i,k) \in \mathcal{E}} \frac{w_k^*}{w_i^* + w_k^*}, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

- Stationary distribution of $\boldsymbol{P}^*$:

$$\boldsymbol{\pi}^* := \frac{1}{\sum_{i=1}^n w_i^*} \left[w_1^*, w_2^*, \ldots, w_n^* \right]^\top$$

Rationale behind spectral method

In large-sample case, $\boldsymbol{P} \xrightarrow{L\to\infty} \boldsymbol{P}^* = [P^*_{i,j}]_{1\le i,j\le n}$:

$$P^*_{i,j} = \begin{cases} \frac{1}{d}\frac{w_j^*}{w_i^*+w_j^*}, & \text{if } (i,j)\in\mathcal{E}, \\ 1-\frac{1}{d}\sum_{k:(i,k)\in\mathcal{E}}\frac{w_k^*}{w_i^*+w_k^*}, & \text{if } i=j, \\ 0, & \text{otherwise.} \end{cases}$$

- Stationary distribution of $\boldsymbol{P}^*$:

$$\boldsymbol{\pi}^* := \frac{1}{\sum_{i=1}^n w_i^*}\left[w_1^*, w_2^*, \ldots, w_n^*\right]^\top$$

- Check detailed balance!

Regularized MLE

Negative log-likelihood

$$\mathcal{L}(\boldsymbol{w}) := -\sum_{(i,j)\in\mathcal{G}} \left\{ y_{j,i} \log \frac{w_i}{w_i + w_j} + (1 - y_{j,i}) \log \frac{w_j}{w_i + w_j} \right\}$$

Regularized MLE

Negative log-likelihood

$$\mathcal{L}(\boldsymbol{w}) := -\sum_{(i,j)\in\mathcal{G}} \left\{ y_{j,i} \log \frac{w_i}{w_i + w_j} + (1 - y_{j,i}) \log \frac{w_j}{w_i + w_j} \right\}$$

$$\xrightarrow{\theta_i = \log w_i} \quad \mathcal{L}(\boldsymbol{\theta}) := \sum_{(i,j)\in\mathcal{G}} \left\{ -y_{j,i} (\theta_i - \theta_j) + \log \left(1 + e^{\theta_i - \theta_j} \right) \right\}$$

Regularized MLE

Negative log-likelihood

$$\mathcal{L}\left(\boldsymbol{w}\right) := -\sum_{(i,j)\in\mathcal{G}}\left\{y_{j,i}\log\frac{w_i}{w_i+w_j} + (1-y_{j,i})\log\frac{w_j}{w_i+w_j}\right\}$$

$$\xrightarrow{\theta_i=\log w_i} \quad \mathcal{L}\left(\boldsymbol{\theta}\right) := \sum_{(i,j)\in\mathcal{G}}\left\{-y_{j,i}\left(\theta_i-\theta_j\right) + \log\left(1+e^{\theta_i-\theta_j}\right)\right\}$$

(Regularized MLE) minimize$_{\boldsymbol{\theta}}$ $\quad \mathcal{L}_\lambda\left(\boldsymbol{\theta}\right) := \mathcal{L}\left(\boldsymbol{\theta}\right) + \frac{1}{2}\lambda\|\boldsymbol{\theta}\|_2^2$

choose $\lambda \asymp \sqrt{\frac{np\log n}{L}}$

Prior art

	mean square error for estimating scores	top-K ranking accuracy	
Spectral method	✔	?	Negahban et al. '12
MLE	✔	?	Negahban et al. '12 Hajek et al. '14
Spectral MLE	✔	✔	Chen & Suh. '15

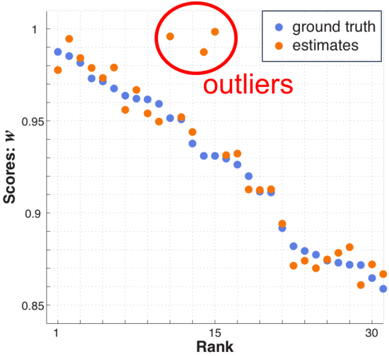
Prior art

“meta metric”

⇧

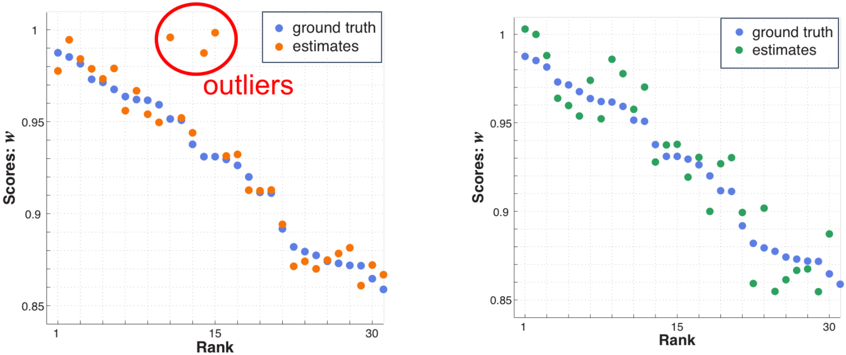
	mean square error for estimating scores	top-K ranking accuracy	
Spectral method	✔	?	Negahban et al. ‘12
MLE	✔	?	Negahban et al. ’12 Hajek et al. ‘14
Spectral MLE	✔	✔	Chen & Suh. ’15

Small ℓ_2 loss $\neq$ high ranking accuracy



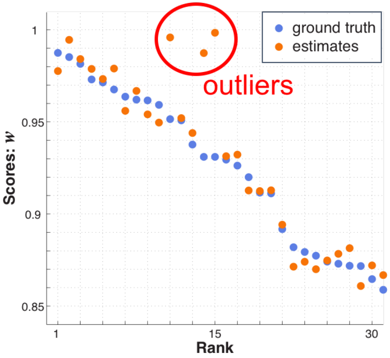
Top 3 : $\{15, 11, 2\}$

Small ℓ_2 loss $\neq$ high ranking accuracy

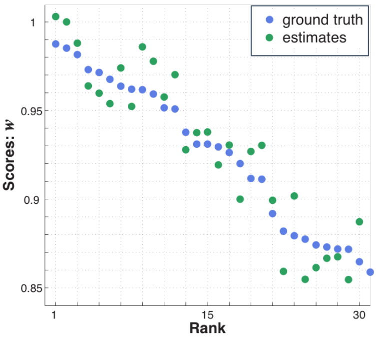


Top 3 : {15, 11, 2}   Top 3: {1, 2, 3}

Small ℓ_2 loss $\neq$ high ranking accuracy



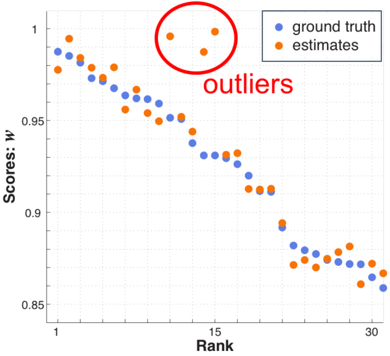
Top 3 : {15, 11, 2}



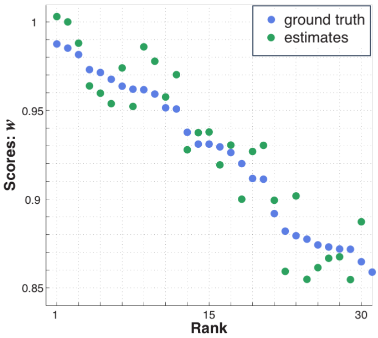
Top 3: {1, 2, 3}

These two estimates have same ℓ_2 loss, but output different rankings

Small ℓ_2 loss $\neq$ high ranking accuracy



Top 3 : $\{15, 11, 2\}$



Top 3: $\{1, 2, 3\}$

These two estimates have same ℓ_2 loss, but output different rankings

Need to control entrywise error!

Optimality?

Is spectral method or MLE alone optimal for top-K ranking?

Optimality?

Is spectral method or MLE alone optimal for top-K ranking?

Partial answer (Jang et al '16):

spectral method works if comparison graph is sufficiently dense

Optimality?

Is spectral method or MLE alone optimal for top-K ranking?

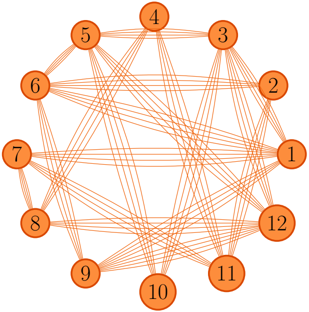
Partial answer (Jang et al '16):

spectral method works if comparison graph is sufficiently dense

This work: affirmative answer for both methods + $\underbrace{\text{entire regime}}_{\text{inc. sparse graphs}}$

Main result

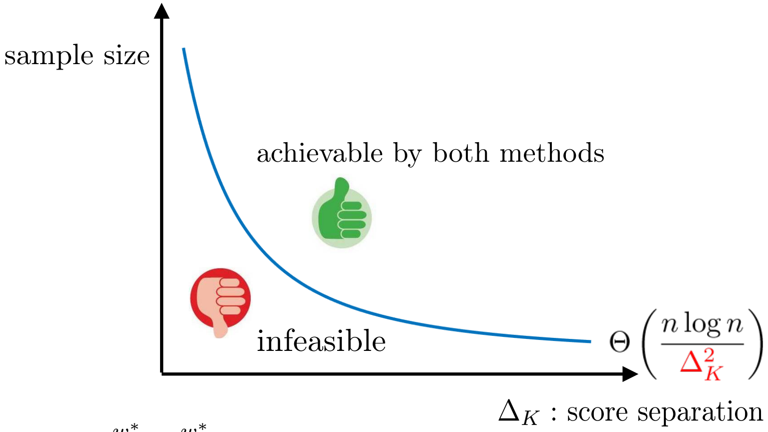
comparison graph $\mathcal{G}(n,p)$; sample size $\asymp pn^2 L$



Theorem 1 (Chen, Fan, Ma, Wang '17)

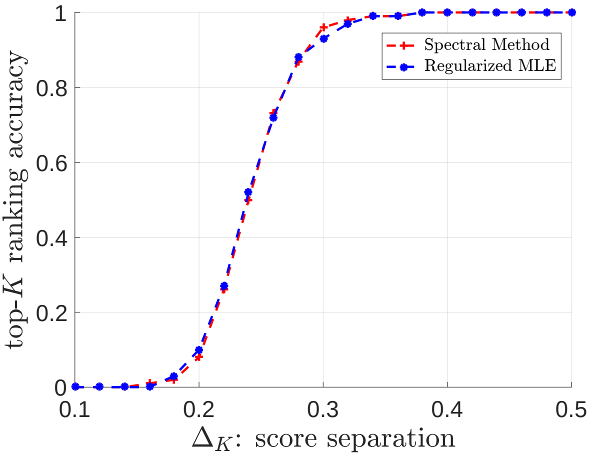
When $p \gtrsim \frac{\log n}{n}$, both spectral method and regularized MLE achieve *optimal sample complexity* *for top-K ranking!*

Main result



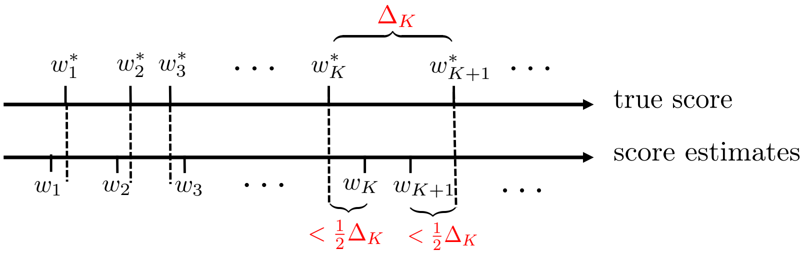
- $\Delta_K := \frac{w^*_{(K)} - w^*_{(K+1)}}{\|\boldsymbol{w}^*\|_\infty}$: score separation

Empirical top-K ranking accuracy



$$n = 200, \ p = 0.25, \ L = 20$$

Optimal control of entrywise error



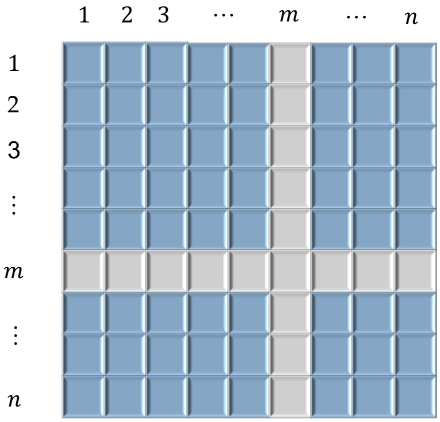
Theorem 2

Suppose $p \gtrsim \frac{\log n}{n}$ *and sample size* $\gtrsim \frac{n \log n}{\Delta_K^2}$. *Then with high prob., the estimates* $\boldsymbol{w}$ *returned by both methods obey (up to global scaling)*

$$\frac{\|\boldsymbol{w} - \boldsymbol{w}^*\|_\infty}{\|\boldsymbol{w}^*\|_\infty} < \frac{1}{2}\Delta_K$$

Key ingredient: leave-one-out analysis

For each $1 \leq m \leq n$, introduce leave-one-out estimate $\boldsymbol{w}^{(m)}$



$$\Longrightarrow \boldsymbol{w}^{(m)}$$

$$\boldsymbol{y} = [y_{i,j}]_{1 \leq i,j \leq n}$$

Leave-one-out stability

leave-one-out estimate $\boldsymbol{w}^{(m)} \approx$ true estimate $\boldsymbol{w}$

Leave-one-out stability

leave-one-out estimate $\boldsymbol{w}^{(m)} \approx$ true estimate $\boldsymbol{w}$

- Spectral method: eigenvector perturbation bound

$$\|\boldsymbol{\pi} - \widehat{\boldsymbol{\pi}}\|_{\boldsymbol{\pi}^*} \lesssim \frac{\|\boldsymbol{\pi}(\boldsymbol{P} - \widehat{\boldsymbol{P}})\|_{\boldsymbol{\pi}^*}}{\text{spectral-gap}}$$

  - new Davis-Kahan bound for $\underbrace{\text{probability transition matrices}}_{\textit{asymmetric}}$

Leave-one-out stability

leave-one-out estimate $\boldsymbol{w}^{(m)}$ $\approx$ true estimate $\boldsymbol{w}$

- Spectral method: eigenvector perturbation bound

$$\|\boldsymbol{\pi} - \widehat{\boldsymbol{\pi}}\|_{\boldsymbol{\pi}^*} \lesssim \frac{\left\|\boldsymbol{\pi}(\boldsymbol{P} - \widehat{\boldsymbol{P}})\right\|_{\boldsymbol{\pi}^*}}{\text{spectral-gap}}$$

  - new Davis-Kahan bound for $\underbrace{\text{probability transition matrices}}_{\textit{asymmetric}}$

- MLE: local strong convexity

$$\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_2 \lesssim \frac{\|\nabla \mathcal{L}_\lambda(\boldsymbol{\theta}; \widehat{\boldsymbol{y}})\|_2}{\text{strong convexity parameter}}$$

Summary

	Optimal sample complexity	Linear-time computational complexity
Spectral method	✔	✔
Regularized MLE	✔	✔

Novel entrywise perturbation analysis for spectral method and convex optimization

Paper: "*Spectral method and regularized MLE are both optimal for top-K ranking*", Y. Chen, J. Fan, C. Ma, K. Wang, arxiv:1707.09971, 2017