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

Yuxin Chen
Electrical Engineering, Princeton University

Joint work with Jianqing Fan, Cong Ma 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 score to each of \( n \) items \( \mathbf{w}^* = [w_1^*, \ldots, w_n^*] \)
Assign latent score to each of $n$ items $\boldsymbol{w}^* = [w_1^*, \cdots, w_n^*]$.

- **This work:** Bradley-Terry-Luce (logistic) model

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

- **Other models:** Thurstone model, low-rank model, ...
Typical ranking procedures

Estimate latent scores

→ rank items based on score estimates
**Top-$K$ ranking**

Estimate latent scores

→ rank items based on score estimates

**Goal:** identify the set of top-$K$ items under minimal sample size
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

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

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

<table>
<thead>
<tr>
<th></th>
<th>mean square error for estimating scores</th>
<th>top-K ranking accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral method</td>
<td>✔</td>
<td>?</td>
</tr>
<tr>
<td>MLE</td>
<td>✔</td>
<td>?</td>
</tr>
<tr>
<td>Spectral MLE</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

- Negahban et al. ‘12
- Negahban et al. ‘12
- Hajek et al. ‘14
- Chen & Suh. ‘15
## Prior art

### “meta metric”

<table>
<thead>
<tr>
<th></th>
<th>Mean square error for estimating scores</th>
<th>Top-K ranking accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral method</td>
<td>✔</td>
<td>?</td>
</tr>
<tr>
<td>MLE</td>
<td>✔</td>
<td>?</td>
</tr>
<tr>
<td>Spectral MLE</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Negahban et al. ‘12
Negahban et al. ‘12
Hajek et al. ‘14
Chen & Suh. ‘15
Small $\ell_2$ loss $\neq$ high ranking accuracy

These two estimates have same $\ell_2$ loss, but output different rankings

Need to control entrywise error!

Top 3: \{15, 11, 2\}
Small $\ell_2$ loss $\neq$ high ranking accuracy

These two estimates have same $\ell_2$ loss, but output different rankings.

Need to control entrywise error!

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

Top 3: $\{1, 2, 3\}$
Small $\ell_2$ loss $\neq$ high ranking accuracy

These two estimates have same $\ell_2$ loss, but output different rankings

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

Top 3: $\{1, 2, 3\}$
Small $\ell_2$ loss $\neq$ high ranking accuracy

These two estimates have same $\ell_2$ loss, but output different rankings

Need to control entrywise error!
Partial answer (Jang et al ’16): spectral method works if comparison graph is sufficiently dense. This work: affirmative answer for both methods + entire regime including sparse graphs.
Is spectral method or MLE alone optimal for top-$K$ ranking?

Partial answer (Jang et al ’16):

\textit{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 + entire regime

inc. sparse graphs
Spectral method (Rank Centrality)

Negahban, Oh, Shah ’12

- Construct a probability transition matrix $P$, whose off-diagonal entries obey

$$P_{i,j} \propto \begin{cases} y_{i,j}, & \text{if } (i, j) \in G \\ 0, & \text{if } (i, j) \notin G \end{cases}$$

- Return score estimate as leading left eigenvector of $P$
Rationale behind spectral method

In large-sample case, \( P \rightarrow P^* \), whose off-diagonal entries obey

\[
P_{i,j}^* \propto \begin{cases} 
\frac{w_j^*}{w_i^* + w_j^*}, & \text{if } (i, j) \in \mathcal{G} \\
0, & \text{if } (i, j) \notin \mathcal{G}
\end{cases}
\]

- Stationary distribution of \( P^* \) is reversible and satisfies detailed balance.

\[
\pi^* \propto [w_1^*, w_2^*, \ldots, w_n^*]
\]

true score
Regularized MLE

Negative log-likelihood

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

- \( \mathcal{L}(w) \) becomes convex after reparametrization:

\[ w \rightarrow \theta = [\theta_1, \ldots, \theta_n], \quad \theta_i = \log w_i \]
Regularized MLE

Negative log-likelihood

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

- $\mathcal{L}(w)$ becomes convex after reparametrization:

$$w \rightarrow \theta = [\theta_1, \cdots, \theta_n], \quad \theta_i = \log w_i$$

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

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

comparison graph $G(n, p)$; sample size $\precsim pn^2L$

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

- Infeasible feasible sample size

Comparison with Jang et al'16
- Jang et al'16: spectral method controls entrywise error if $p \leq \log n \cdot \sqrt[4]{\varepsilon}$ relatively dense
- Our work / optimal sample size $K \cdot \log n \cdot \frac{1}{4}$

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

$\Theta \left( \frac{n \log n}{\Delta^2_K} \right)$

$\Delta_K$: score separation

Top-$K$ ranking
Comparison with Jang et al ’16

Jang et al ’16: spectral method controls entrywise error if \( p \gtrsim \sqrt{\frac{\log n}{n}} \) relatively dense

Top-\( K \) ranking
Comparison with Jang et al ’16

Jang et al ’16: spectral method controls entrywise error if \( p \gtrsim \sqrt{\frac{\log n}{n}} \) (relatively dense)

\[
K \downarrow \log n \quad n \uparrow \frac{1}{4}
\]

Our work / optimal sample size

\[
K \downarrow \log n \quad n \uparrow \frac{1}{4}
\]

score separation: \( K : \) score separation

Top-\( K \) ranking
Empirical top-$K$ ranking accuracy

$n = 200$, $p = 0.25$, $L = 20$
Theorem 2

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

$$\frac{\|\mathbf{w} - \mathbf{w}^*\|_{\infty}}{\|\mathbf{w}^*\|_{\infty}} < \frac{1}{2} \Delta_K$$
Key ingredient: leave-one-out analysis

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

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

$$|w_m - w^*_m| \leq \underbrace{|w_{m}^{(m)} - w^*_m|}_{\text{Leave-one-out estimation error}} + \underbrace{\|w - w^{(m)}\|_2}_{\text{Leave-one-out perturbation}}$$

statistical independence

stability
Exploit statistical independence

Key ingredient: leave-one-out analysis

For each $m \in [n]$, introduce leave-one-out estimate $w^{(m)}$

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

leave-one-out estimate $w^{(m)}$ \(\perp\) all data related to $m$th item
Leave-one-out stability

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

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

- Spectral method: eigenvector perturbation bound

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

- new Davis-Kahan bound for probability transition matrices

\textit{asymmetric}
Leave-one-out stability

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

- Spectral method: eigenvector perturbation bound

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

  - new Davis-Kahan bound for probability transition matrices

- MLE: local strong convexity

$$\| \theta - \hat{\theta} \|_2 \lesssim \frac{\| \nabla \mathcal{L}_\lambda (\theta; \hat{y}) \|_2}{\text{strong convexity parameter}}$$
A small sample of related works

• Parametric models
  o Ford ’57
  o Hunter ’04
  o Negahban, Oh, Shah ’12
  o Rajkumar, Agarwal ’14
  o Hajek, Oh, Xu ’14
  o Chen, Suh ’15
  o Rajkumar, Agarwal ’16
  o Jang, Kim, Suh, Oh ’16
  o Suh, Tan, Zhao ’17

• Non-parametric models
  o Shah, Wainwright ’15
  o Shah, Balakrishnan, Guntuboyina, Wainwright ’16
  o Chen, Gopi, Mao, Schneider ’17

• Leave-one-out analysis
  o El Karoui, Bean, Bickel, Lim, Yu ’13
  o Zhong, Boumal ’17
  o Abbe, Fan, Wang, Zhong ’17
### Summary

<table>
<thead>
<tr>
<th></th>
<th>Optimal sample complexity</th>
<th>Linear-time computational complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral method</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Regularized MLE</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

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