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

Top-$K$ ranking
Parametric models

Assign **latent preference score** to each of $n$ items $\mathbf{w}^* = [w^*_1, \ldots, w^*_n]$
Assign latent preference score to each of \( n \) items \( \mathbf{w}^* = [w_1^*, \ldots, w_n^*] \)

- **This work:** Bradley-Terry-Luce model: for \( \mathbf{w}^* \in \mathbb{R}_+^n \)
  \[
  \mathbb{P} \{ \text{item } j \text{ beats item } i \} = \frac{w_j^*}{w_i^* + w_j^*}
  \]
Other parametric models

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

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

Gaussian cdf
Other parametric models

• Thurstone model: for \( \mathbf{w}^* \in \mathbb{R}^n \)

\[
P\{\text{item } j \text{ beats item } i\} = \Phi \left( \mathbf{w}_j^* - \mathbf{w}_i^* \right)
\]

Gaussian cdf

• Parametric models: for nondecreasing \( f : \mathbb{R} \to [0, 1] \) which obey

\[
f(t) = 1 - f(-t), \quad \forall t \in \mathbb{R}
\]

Then we set

\[
P\{\text{item } j \text{ beats item } i\} = f \left( \mathbf{w}_j^* - \mathbf{w}_i^* \right)
\]
Typical ranking procedures

Estimate latent scores
  \[\rightarrow\] rank items based on score estimates
Top-$K$ ranking

Estimate latent scores

→ rank items based on score estimates

Goal: identify the set of \textbf{top-$K$ items} with pairwise comparisons
Model: random sampling

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

- For each $(i, j) \in G$, 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 $G \sim \mathcal{G}(n, p)$

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

$$ y_{i,j} = \frac{1}{L} \sum_{l=1}^{L} y_{i,j}^{(l)} \quad \text{(sufficient statistic)} $$
### Spectral method (Rank Centrality)

Negahban, Oh, Shah ’12

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

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

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

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

\[
P^*_{i,j} = \begin{cases} 
\frac{1}{d} \frac{w^*_j}{w^*_i + w^*_j}, & \text{if } (i, j) \in \mathcal{E}, \\
1 - \frac{1}{d} \sum_{k : (i, k) \in \mathcal{E}} \frac{w^*_k}{w^*_i + w^*_k}, & \text{if } i = j, \\
0, & \text{otherwise.}
\end{cases}
\]
Rationale behind spectral method

In large-sample case, \( \mathbf{P} \xrightarrow{L \to \infty} \mathbf{P}^* = [P^*_{i,j}]_{1 \leq i,j \leq n} \):

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

• Stationary distribution of \( \mathbf{P}^* \):

\[
\pi^* := \frac{1}{\sum_{i=1}^{n} w_i^*} [w_1^*, w_2^*, \ldots, w_n^*]^T
\]
Rationale behind spectral method

In large-sample case, \( P \xrightarrow{L \to \infty} P^* = [P_{i,j}^*]_{1 \leq i,j \leq n} \):

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

- Stationary distribution of \( P^* \):

\[
\pi^* := \frac{1}{\sum_{i=1}^{n} w_i^*} [w_1^*, w_2^*, \ldots, w_n^*]^T
\]

- Check detailed balance!
Regularized MLE

Negative log-likelihood

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

Negative log-likelihood

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

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

Regularized MLE

Negative log-likelihood

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

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

\begin{align*}
\text{(Regularized MLE)} \quad \text{minimize}_{\theta} \quad & \mathcal{L}_\lambda(\theta) := \mathcal{L}(\theta) + \frac{1}{2} \lambda \| \theta \|^2_2 \\
\text{choose} \quad & \lambda \approx \sqrt{\frac{np \log n}{L}}
\end{align*}
## Prior art

<table>
<thead>
<tr>
<th>Method</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

Hajek et al. ‘14

Chen & Suh. ‘15
### Prior art

<table>
<thead>
<tr>
<th>“meta metric”</th>
<th>mean square error for estimating scores</th>
<th>top-K ranking accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-K ranking</td>
<td></td>
<td>11/20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Spectral method</th>
<th>MLE</th>
<th>Spectral MLE</th>
<th>Negahban et al. ‘12</th>
<th>Negahban et al. ‘12</th>
<th>Hajek et al. ‘14</th>
<th>Chen &amp; Suh. ‘15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Small $\ell_2$ loss $\neq$ high ranking accuracy

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

Need to control entrywise error!

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

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
**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!
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 + entire regime inc. sparse graphs
Main result

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

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!

Top-$K$ ranking
Main result

\[ \Delta_K := \frac{w^*_K - w^*_{K+1}}{\|w^*\|_\infty} : \text{score separation} \]
Empirical top-$K$ ranking accuracy

$n = 200, \ p = 0.25, \ L = 20$
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 \( \mathbf{w} \) 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 $\mathbf{w}^{(m)}$

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

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

- new Davis-Kahan bound for probability transition matrices

\textit{asymmetric}
Leave-one-out stability

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

• 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

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