A Spectral Algorithm for Ranking using Seriation

Alex d’Aspremont, CNRS & ENS, Paris.

with Fajwel Fogel, Milan Vojnovic,
Ecole Polytechnique, Microsoft Research.
Given \( n \) items, and \textbf{pairwise comparisons} \( \text{item}_i \succ \text{item}_j \), for \((i, j) \in S\),

find a global \textbf{ranking} \( \pi(i) \) of these items

\[
\text{item}_{\pi(1)} \succ \text{item}_{\pi(2)} \succ \ldots \succ \text{item}_{\pi(n)}
\]
Pairwise comparisons?

- Some data sets naturally produce pairwise comparisons, e.g. tournaments, ecommerce transactions, etc.
- Comparing items is often more intuitive than ranking them directly.

Hot or Not? Rank images by "hotness"...
Ranking & pairwise comparisons

[Jamieson and Nowak, 2011]
Classical problem, many algorithms \textit{(roughly sorted by increasing complexity)}

- **Scores.** Borda, Elo rating system (chess), TrueSkill [Herbrich et al., 2006], etc.
- **Spectral methods.** [Dwork et al., 2001, Negahban et al., 2012]
- **MLE based algorithms.** [Bradley and Terry, 1952, Luce, 1959, Herbrich et al., 2006]
- **Learning to rank.** Learn scoring functions.

See forthcoming book by Milan Vojnovic on the subject...
Ranking & pairwise comparisons

Various data settings.

- A partial subset of preferences is observed.
- Preferences are measured actively [Ailon, 2011, Jamieson and Nowak, 2011].
- Preferences are fully observed but arbitrarily corrupted.
- Repeated noisy observations.

Various performance metrics.

- Minimize the number of disagreements i.e. $\#$ edges inconsistent with the global ordering, e.g. PTAS for min. Feedback Arc Set (FAS) [Kenyon-Mathieu and Schudy, 2007]
- Maximize likelihood given a model on pairwise observations, e.g. [Bradley and Terry, 1952, Luce, 1959]
- Convex loss in ranking SVMs
- ...
Scores

- **Borda count.** Dates back at least to 18th century. Players are ranked according to the number of wins divided by the total number of comparisons [Ammar and Shah., 2011, Wauthier et al., 2013].

- Also **fair-bets, invariant scores, Masey, Maas, Colley**, etc.

- **Elo rating system.** (Chess, \(\sim 1970\))
  
  - Players have skill levels \(\mu_i\).
  - Probability of player \(i\) beating \(j\) given by \(H(\mu_i - \mu_j)\) (e.g. Gaussian CDF).
  - After each game, skills updated to

\[
\mu_i := \mu_i + c(w_{ij} - H(\mu_i - \mu_j))
\]

\(w_{ij} \in \{0, 1\}\), \(H(\cdot)\) e.g. Gaussian CDF. Sum of all skills remains constant, skills are transferred from losing players to winning ones.

- **TrueSkill rating system.** [Herbrich et al., 2006] Similar in spirit to Elo, but player skills are represented by a Gaussian distribution.

Very low numerical cost.
Spectral algorithms

**Spectral algorithms**: Markov chain on a graph.

Similar to HITS [Kleinberg, 1999] or Pagerank [Page et al., 1998]. . .

- A **random walk** goes through a graph where each node corresponds to an item or player to rank.
- Likelihood of going from i to j depends on how often i lost to j, so neighbors with more wins will be visited more frequently.
- Ranking is based on asymptotic frequency of visits, i.e. on the **stationary distribution**.

Simple extremal eigenvalue computation, low complexity (roughly $O(n^2 \log n)$ in the dense case). See Dwork et al. [2001], Negahban et al. [2012] for more details.
Model pairwise comparisons. [Bradley and Terry, 1952, Luce, 1959, Herbrich et al., 2006]

- Comparisons are generated according to a generalized linear model (GLM).
- Repeated observations, independent. Item \( i \) is preferred over item \( j \) with probability
  \[
  P_{i,j} = H(\nu_i - \nu_j)
  \]
  where
  - \( \nu \in \mathbb{R}^n \) is a vector of skills.
  - \( H: \mathbb{R} \to [0, 1] \) is an increasing function.
  - \( H(-x) = 1 - H(x) \), and \( \lim_{x \to -\infty} H(x) = 0 \) and \( \lim_{x \to \infty} H(x) = 1 \).

\( H(\cdot) \) is a CDF. Logistic in the Bradley-Terry-Luce model: \( H(x) = 1/(1 + e^{-x}) \).
Also: Gaussian (Thurstone), Laplace, etc.

Estimate \( \nu \) by maximizing likelihood. (using e.g. fixed point algo)
Current tradeoffs.

- Scoring methods are exact in the noiseless case, but not very robust.
- Spectral methods are more robust, but no exact recovery guarantee (error bounds in BTL by [Negahban et al., 2012]).
- Learning to rank methods are very expensive.

Today

- Connect ranking from **pairwise preferences** to ranking based on **similarity**.
- Solution also given by spectral algorithm, but completely different from the “Markov chain on a graph” argument.
Outline

- Introduction
- **Seriation**
  - From ranking to seriation
  - Robustness
- Numerical results
Seriation

The Seriation Problem.

- Pairwise similarity information $A_{ij}$ on $n$ variables.
- Suppose the data has a serial structure, i.e. there is an order $\pi$ such that

$$A_{\pi(i)\pi(j)} \text{ decreases with } |i - j| \quad (R\text{-matrix})$$

Recover $\pi$?

![Similarity matrix](image1)
![Input](image2)
![Reconstructed](image3)

Alex d’Aspremont

Simons Institute, Berkeley, September 2014, 12/35
**A Spectral Solution**

**Spectral Clustering.** Define the Laplacian of $A$ as $L_A = \text{diag}(A1) - A$, the Fiedler vector of $A$ is written

$$f = \arg\min_{1^T x = 0, \|x\|_2 = 1} x^T L_A x.$$ 

and is the second smallest eigenvector of the Laplacian.

The Fiedler vector reorders a $R$-matrix in the noiseless case.

---

**Theorem [Atkins, Boman, Hendrickson, et al., 1998]**

**Spectral seriation.** Suppose $A \in \mathbb{S}_n$ is a pre-$R$ matrix, with a simple Fiedler value whose Fiedler vector $f$ has no repeated values. Suppose that $\Pi \in \mathcal{P}$ is such that the permuted Fiedler vector $\Pi v$ is monotonic, then $\Pi A \Pi^T$ is an $R$-matrix.
C1P has direct applications in shotgun gene sequencing.

- Genomes are cloned multiple times and randomly cut into shorter reads (∼ 400bp), which are fully sequenced.
- Reorder the reads to recover the genome.

(from Wikipedia. . . )
Outline

- Introduction
- Seriation
- From ranking to seriation
- Robustness
- Numerical results
From Ranking to Seriation

Similarity matrices from pairwise comparisons.

- Given pairwise comparisons \( C \in \{-1, 0, 1\}^{n \times n} \) with

  \[
  C_{i,j} = \begin{cases} 
  1 & \text{if } i \text{ is ranked higher than } j \\
  0 & \text{if } i \text{ and } j \text{ are not compared or in a draw} \\
  -1 & \text{if } j \text{ is ranked higher than } i
  \end{cases}
  \]

- Define the pairwise similarity matrix \( S_{\text{match}} \) as

  \[
  S_{i,j}^{\text{match}} = \sum_{k=1}^{n} \left( \frac{1 + C_{i,k}C_{j,k}}{2} \right).
  \]

  \( S_{i,j}^{\text{match}} \) counts the number of matching comparisons between \( i \) and \( j \) with other reference items \( k \).

In a tournament setting: players that beat the same players and are beaten by the same players should have a similar ranking....
Similarity from preferences. Given all comparisons $C_{i,j} \in \{-1, 0, 1\}$ between items ranked linearly, the similarity matrix $S^{\text{match}}$ is a strict R-matrix and

$$S_{ij}^{\text{match}} = n - |i - j|$$

for all $i, j = 1, \ldots, n$.

This means that, given all pairwise comparisons, spectral clustering on $S^{\text{match}}$ will recover the true ranking.
Similarity matrices in the generalized linear model.

- Observations are independent, and item $i$ is preferred over item $j$ with probability
  $$P_{i,j} = H(\nu_i - \nu_j)$$
  with $H(\cdot)$ CDF.
- We estimate the matrix
  $$S_{i,j}^{\text{glm}} \to \sum_{k=1}^{n} \left( 1 - \frac{|P_{i,k} - P_{j,k}|}{2} \right).$$
  based on (repeated) preference observations.
From Ranking to Seriation

[Fogel et al., 2014]

**Similarity from preferences in the GLM.** *If the items are ordered by decreasing values of the skill parameters, the similarity matrix $S^{glm}$ is a strict R matrix.*

This means that, given enough samples on pairwise comparisons, spectral clustering on $S^{glm}$ will recover the true ranking.
Outline

- Introduction
- Seriation
- From ranking to seriation
- Robustness
- Numerical results
Robustness to corrupted entries.

- Given all comparisons $C_{s,t} \in \{-1, 1\}$ between items ordered $1, \ldots, n$.
- Suppose the sign of one comparison $C_{i,j}$ is switched, with $i < j$.

If $j - i > 2$ then $S_{\text{match}}$ remains a strict-$R$ matrix.

In this case, the score vector $w$ has ties between items $i$ and $i + 1$ and items $j$ and $j - 1$. 
Robustness

A graphical argument...

The matrix of pairwise comparisons $C$ (far left).

The corresponding similarity matrix $S^{\text{match}}$ is a strict R-matrix (center left).

The same $S^{\text{match}}$ similarity matrix with comparison (3,8) corrupted (center right). With one corrupted comparison, $S^{\text{match}}$ keeps enough strict R-constraints to recover the right permutation. (far right).
Robustness

Generalizes to several errors. . .

[Fogel et al., 2014]

**Robustness to corrupted entries.** Given a comparison matrix for a set of \( n \) items with \( m \) corrupted comparisons selected uniformly at random from the set of all possible item pairs. The probability of recovery \( p(n, m) \) using seriation on \( S_{\text{match}} \) satisfies \( p(n, m) \geq 1 - \delta \), provided that \( m = O(\sqrt{\delta n}) \).

- One corrupted comparison is enough to create ambiguity in scoring arguments.
- Need \( \Omega(n^2) \) comparisons for exact recovery [Jamieson and Nowak, 2011].
- No exact recovery results for Markov Chain type spectral methods.
Robustness

We can go bit further. . . .

Form $S^{\text{match}}$ from consistent, ordered comparisons.

Much simpler to analyze than MC methods: using results from [Von Luxburg et al., 2008], we can compute its Fiedler vector asymptotically.

The Fiedler vector of the nonsymmetric normalized Laplacian is also given by $x_i = c i$, $i = 1, \ldots, n$ where $c > 0$, for finite $n$.

The spectral gap between the first three eigenvalues can be controlled.
Robustness

- Asymptotically: $S_{\text{match}} / n \to k(x, y) = 1 - |x - y|$ for $x, y \in [0, 1]$.
- The degree function is then $d(x) = \int_0^1 k(x, y) dy = -x^2 + x + 1/2$.
- The range of $d(x)$ is $[0.5, 0.75]$ and the bulk of the spectrum is contained in this interval.
- The Fiedler vector $f$ with eigenvalue $\lambda$ satisfies
  \[
  f''(x)(1/2 - \lambda + x - x^2) + 2f'(x)(1 - 2x) = 0.
  \]
- We can also show that the second smallest eigenvalues of the unnormalized Laplacian satisfies $\lambda_2 < 2/5$, which is outside of this range.

Von Luxburg et al. [2008] then show that the unnormalized Laplacian converges and that its second eigenvalue is simple. Idem for the normalized Laplacian.

This spectral gap means we can use **perturbation analysis** to study recovery.
Robustness

- **Perturbation analysis** shows that

\[ \|f - \hat{f}\|_2 \leq \sqrt{2} \frac{\|L - \hat{L}\|_2}{\min\{\lambda_2 - \lambda_1, \lambda_3 - \lambda_2\}} \]

where \(L, f\) are the true Laplacian (resp. Fiedler vector) and \(\hat{L}, \hat{f}\) the perturbed ones.

- In fact, we have

\[ \hat{f} = f - R_2Ef + o(\|E\|_2), \quad \text{with } E = (L - \hat{L}) \]

where \(R_2\) is the resolvent

\[ R_2 = \sum_{j \neq 2} \frac{1}{\lambda_j - \lambda_2} u_j u_j^T, \]

- If \(\|f - \hat{f}\|_\infty\) is **smaller than the gap between coefficients** in the leading eigenvector, ranking recovery remains exact.
Robustness

With missing observations, C is **subsampled**, which means that the error $E$ can be controlled as in Achlioptas and McSherry [2007].

- Take a symmetric matrix $M \in \mathbb{S}_n$ whose entries $M$ are independently sampled as

  \[
  S_{ij} = \begin{cases} 
  M_{ij}/p & \text{with probability } p \\
  0 & \text{otherwise},
  \end{cases}
  \]

  where $p \in [0, 1]$.

- Theorem 1.4 in Achlioptas and McSherry [2007] shows that when $n$ is large enough

  \[
  \|M - S\|_2 \leq 4\|M\|_\infty \sqrt{n/p},
  \]

  holds with high probability.
Comparing the asymptotic Fiedler vector, and the true one for $n = 100$. 
Outline

- Introduction
- Seriation
- From ranking to seriation
- Robustness
- Numerical results
Uniform noise/corruption. Kendall $\tau$ (higher is better) for **SerialRank** (SR, full red line), **row-sum** (PS, [Wauthier et al., 2013] dashed blue line), **rank centrality** (RC [Negahban et al., 2012] dashed green line), and **maximum likelihood** (BTL [Bradley and Terry, 1952], dashed magenta line).
Percentage of upsets (i.e. disagreeing comparisons, lower is better), for various values of $k$ and ranking methods, on TopCoder (left) and football data (right).
Percentage of upsets (i.e. disagreeing comparisons, lower is better), for various values of $k$ and ranking methods, on England Premier League 2011-2012 season (left) and 2012-2013 season (right).
<table>
<thead>
<tr>
<th>Official</th>
<th>Row-sum</th>
<th>RC</th>
<th>BTL</th>
<th>SerialRank</th>
<th>Semi-Supervised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man City (86)</td>
<td>Man City</td>
<td>Liverpool</td>
<td>Man City</td>
<td>Man City</td>
<td>Man City</td>
</tr>
<tr>
<td>Liverpool (84)</td>
<td>Liverpool</td>
<td>Arsenal</td>
<td>Liverpool</td>
<td>Liverpool</td>
<td>Chelsea</td>
</tr>
<tr>
<td>Chelsea (82)</td>
<td>Chelsea</td>
<td>Man City</td>
<td>Chelsea</td>
<td>Chelsea</td>
<td>Liverpool</td>
</tr>
<tr>
<td>Arsenal (79)</td>
<td>Arsenal</td>
<td>Chelsea</td>
<td>Arsenal</td>
<td>Arsenal</td>
<td>Liverpool</td>
</tr>
<tr>
<td>Everton (72)</td>
<td>Everton</td>
<td>Everton</td>
<td>Everton</td>
<td>Everton</td>
<td>Everton</td>
</tr>
<tr>
<td>Tottenham (69)</td>
<td>Tottenham</td>
<td>Tottenham</td>
<td>Tottenham</td>
<td>Tottenham</td>
<td>Everton</td>
</tr>
<tr>
<td>Man United (64)</td>
<td>Man United</td>
<td>Man United</td>
<td>Man United</td>
<td>Man United</td>
<td>Everton</td>
</tr>
<tr>
<td>Southampton (56)</td>
<td>Southampton</td>
<td>Southampton</td>
<td>Southampton</td>
<td>Southampton</td>
<td>Man United</td>
</tr>
<tr>
<td>Stoke (50)</td>
<td>Stoke</td>
<td>Stoke</td>
<td>Stoke</td>
<td>Stoke</td>
<td>Southmton</td>
</tr>
<tr>
<td>Newcastle (49)</td>
<td>Newcastle</td>
<td>Newcastle</td>
<td>Newcastle</td>
<td>Newcastle</td>
<td>Newcastle</td>
</tr>
<tr>
<td>Crystal Palace (45)</td>
<td>Crystal Palace</td>
<td>Crystal Palace</td>
<td>Crystal Palace</td>
<td>Crystal Palace</td>
<td>Crystal Palace</td>
</tr>
<tr>
<td>Swansea (42)</td>
<td>Swansea</td>
<td>West Ham</td>
<td>West Brom</td>
<td>West Brom</td>
<td>West Brom</td>
</tr>
<tr>
<td>West Ham (40)</td>
<td>West Ham</td>
<td>Aston Villa</td>
<td>Aston Villa</td>
<td>Aston Villa</td>
<td>West Ham</td>
</tr>
<tr>
<td>Aston Villa (38)</td>
<td>Aston Villa</td>
<td>Sunderland</td>
<td>Sunderland</td>
<td>Sunderland</td>
<td>Cardiff</td>
</tr>
<tr>
<td>Sunderland (38)</td>
<td>Sunderland</td>
<td>West Brom</td>
<td>West Brom</td>
<td>West Brom</td>
<td>Crystal Palace</td>
</tr>
<tr>
<td>Hull (37)</td>
<td>Hull</td>
<td>Hull</td>
<td>Sunderland</td>
<td>Hull</td>
<td>Hull</td>
</tr>
<tr>
<td>West Brom (36)</td>
<td>Norwich</td>
<td>Norwich</td>
<td>Fulham</td>
<td>Norwich</td>
<td>Fulham</td>
</tr>
<tr>
<td>Norwich (33)</td>
<td>Fulham</td>
<td>Norwich</td>
<td>Fulham</td>
<td>Norwich</td>
<td>Norich</td>
</tr>
<tr>
<td>Fulham (32)</td>
<td>Cardiff</td>
<td>Cardiff</td>
<td>Cardiff</td>
<td>Cardiff</td>
<td>Sunderland</td>
</tr>
<tr>
<td>Cardiff (30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cardiff</td>
</tr>
</tbody>
</table>

Jobs

Recruiting postdocs (one or two years) at Ecole Normale Superieure in Paris.

Send your resume to aspremon@di.ens.fr.
Conclusion

Very diverse set of algorithmic solutions.

- Here: new class of spectral methods based on seriation results.
- Exact recovery results are easy to derive.
- Almost completely explicit perturbation analysis.
- More robust in certain settings.

Coming soon.

- Kendall $\tau$ type bounds on approximate recovery.
- Better characterize errors with close to $O(n \log n)$ observations.

References


