Ranking from pairwise comparisons using Seriation

Alex d'Aspremont, *CNRS & ENS, Paris.*

with Fajwel Fogel (Ecole Polytechnique), Milan Vojnovic, (MSR Cambridge).

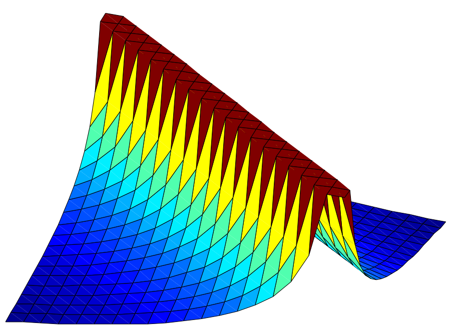
Seriation

The Seriation Problem.

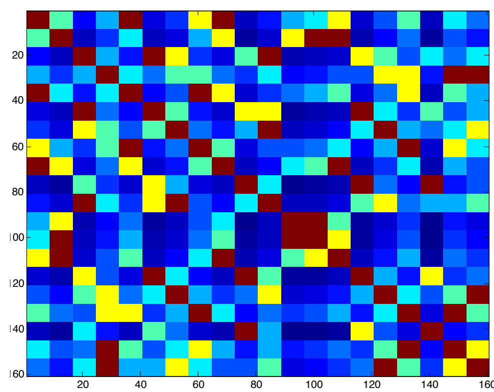
- Pairwise **similarity information** A_{ij} on n variables.
- Suppose the data has a **serial structure**, i.e. there is an order π such that

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

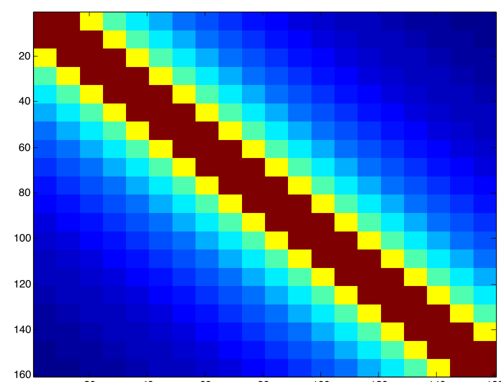
Recover π?



Similarity matrix



Input

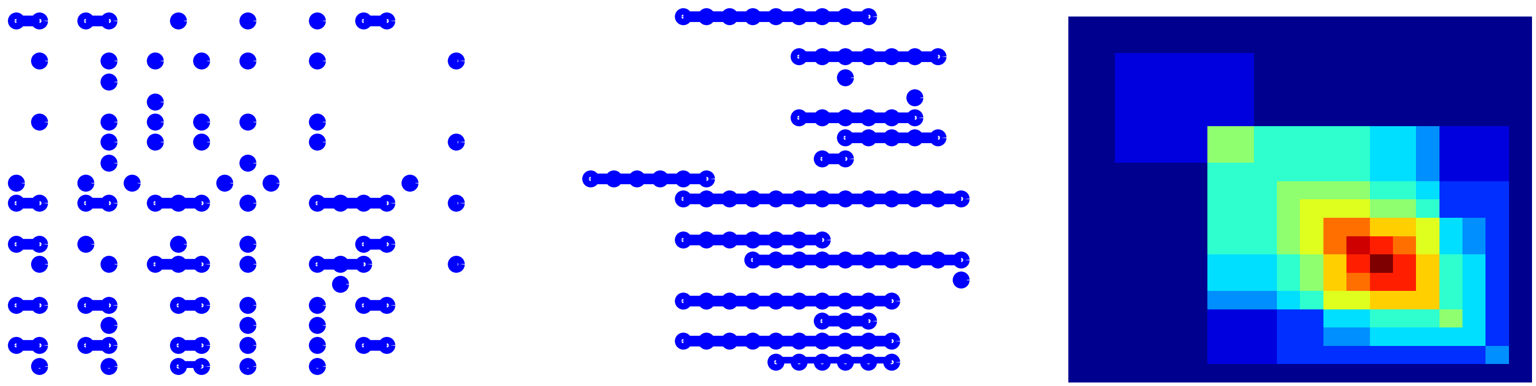


Reconstructed

Seriation

The Continuous Ones Problem.

- We're given a rectangular binary $\{0, 1\}$ matrix.
- Can we reorder its columns so that the ones in each row are contiguous (C1P)?



Input matrix — Ordered C1P matrix — $C^T C$ (overlap)

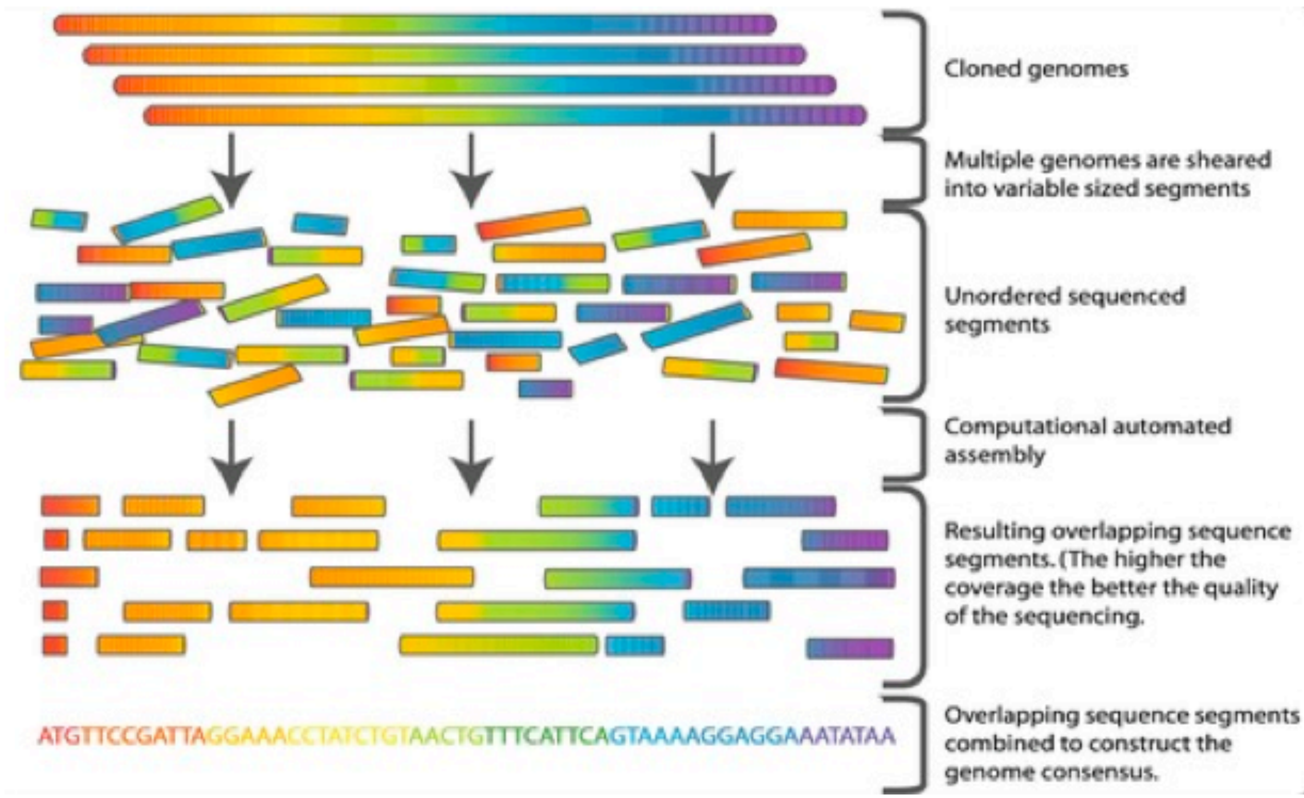
Lemma [Kendall, 1969]

Seriation and C1P. *Suppose there exists a permutation such that C is C1P, then $C\Pi$ is C1P if and only if $\Pi^T C^T C \Pi$ is an R-matrix.*

Shotgun Gene Sequencing

C1P has direct applications in shotgun gene sequencing.

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



(from Wikipedia. . .)

Outline

- Introduction
- **Seriation and 2-SUM**
- Ranking from pairwise comparisons
- Numerical experiments

A Spectral Solution

Spectral Seriation. Define the Laplacian of A as $L_A = \mathbf{diag}(A\mathbf{1}) - A$, the Fiedler vector of A is written

$$f = \underset{\substack{\mathbf{1}^T x = 0,\\ \|x\|_2 = 1}}{\operatorname{argmin}} \; x^T L_A x.$$

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 $A \in \mathbf{S}_n$ 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 Πv is monotonic, then $\Pi A \Pi^T$ is an R-matrix.*

Spectral Solution

A solution in search of a problem. . .

- What if the data is **noisy** and outside the perturbation regime? The spectral solution is only stable when the noise $\|\Delta L\|_2 \leq (\lambda_2 - \lambda_3)/2$.

- What if we have additional **structural information**?

Write seriation as an **optimization problem?**

Seriation and 2-SUM

Combinatorial Solution. Solving 2-SUM

$$\min_{\pi \in \mathcal{P}} \sum_{i,j=1}^{n} A_{\pi(i),\pi(j)}(i-j)^2 = \pi^T L_A \pi \tag{1}$$

and A is a conic combination of CUT (one flat block) matrices.

Laplacian operator is linear, y_π monotonic **optimal for all CUT components.**

Proposition [Fogel et al., 2013]

Seriation and 2-SUM. *Suppose* $C \in \mathbf{S}_n$ *is a* $\{0,1\}$ *pre-R matrix and* $y_i = i$ *for* $i = 1, \ldots, n$. *If* Π *is such that* $\Pi C \Pi^T$ *is an R-matrix, then the permutation* π *solves the 2-SUM combinatorial minimization problem* (1) *for* $A = C^2$.

Recently extended by Laurent and Seminaroti [2014] to the product of R and anti-R Toeplitz matrices.

Convex Relaxation

- Let $\mathcal{D}_n$ the set of doubly stochastic matrices, where

$$\mathcal{D}_n = \{X \in \mathbb{R}^{n \times n} : X \geqslant 0, X\mathbf{1} = \mathbf{1}, X^T\mathbf{1} = \mathbf{1}\}$$

  is the **convex hull of the set of permutation matrices.**

- Notice that $\mathcal{P} = \mathcal{D} \cap \mathcal{O}$, i.e. Π permutation matrix if and only Π is both **doubly stochastic** and **orthogonal.**

Solve

$$\begin{array}{ll}\mbox{minimize} & \mathbf{Tr}(Y^T\Pi^T L_A \Pi Y) - \mu\|P\Pi\|_F^2\\ \mbox{subject to} & e_1^T\Pi g + 1 \leq e_n^T\Pi g,\\ & \Pi\mathbf{1} = \mathbf{1},\ \Pi^T\mathbf{1} = \mathbf{1},\\ & \Pi \geq 0,\end{array} \tag{2}$$

in the variable $\Pi \in \mathbb{R}^{n \times n}$, where $P = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ and $Y \in \mathbb{R}^{n \times p}$ is a matrix whose columns are small perturbations of $g = (1, \ldots, n)^T$.

Convex Relaxation

Other relaxations.

- A lot of work on relaxations for orthogonality constraints, e.g. SDPs in [Nemirovski, 2007, Coifman et al., 2008, So, 2011].
- Simple idea: $Q^T Q = \mathbf{I}$ is a quadratic constraint on Q, **lift it.** This yields a $O(\sqrt{n})$ approximation ratio.
- We could also use $O(\sqrt{\log n})$ approximation bounds for MLA [Even et al., 2000, Feige, 2000, Blum et al., 2000, Rao and Richa, 2005, Feige and Lee, 2007, Charikar et al., 2010].
- All these relaxations form extremely large SDPs.

Our simplest relaxation is a QP. No approximation bounds at this point however.

Semi-Supervised Seriation

Convex Relaxation.

- **Semi-Supervised Seriation.** We can add structural constraints to the relaxation, where

$$a \leq \pi(i) - \pi(j) \leq b \quad \text{is written} \quad a \leq e_i^T \Pi g - e_j^T \Pi g \leq b.$$

  which are linear constraints in Π.

- **Sampling permutations.** We can generate permutations from a doubly stochastic matrix D
  - Sample monotonic random vectors u.
  - Recover a permutation by reordering Du.

- **Algorithms.** Large QP. [Lim and Wright, 2014] recently used results by [Goemans, 2014] on extended formulations of the permutahedron to significantly reduce problem dimension, from $O(n^2)$ to $O(n \log^2 n)$.

Outline

- Introduction
- Seriation and 2-SUM
- **Ranking from pairwise comparisons**
- Numerical experiments

Ranking & pairwise comparisons

Given n items, and **pairwise comparisons**

$$\mathrm{item}_i \succ \mathrm{item}_j, \quad \text{for } (i,j) \in S,$$

find a global **ranking** $\pi(i)$ of these items

$$\mathrm{item}_{\pi(\mathbf{1})} \succ \mathrm{item}_{\pi(\mathbf{2})} \succ \ldots \succ \mathrm{item}_{\pi(\mathbf{n})}$$

Ranking & pairwise comparisons

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

Classical problem, many algorithms *(roughly sorted by increasing complexity)*

- **Scores.** Borda, Elo rating system (chess), TrueSkill [Herbrich et al., 2006], etc.
- **Spectral methods.** [Saaty, 1977, 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 μ_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.

Maximum Likelihood Estimation

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 ν by **maximizing likelihood.** (using e.g. fixed point algo)

Today: Ranking using Seriation

- Produce **similarity matrices** from **pairwise preferences**
- Use **seriation** to recover the ranking.

A **new class of spectral algorithms** for ranking.

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^{match} as

$$S^{\mathrm{match}}_{i,j} = \sum_{k=1}^{n} \left(\frac{1 + C_{i,k} C_{j,k}}{2} \right).$$

- $S^{\mathrm{match}}_{i,j}$ 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. . .

From Ranking to Seriation

[Fogel et al., 2014]

Similarity from preferences. *Given all comparisons $C_{i,j} \in \{-1, 0, 1\}$ between items ranked linearly, the similarity matrix S^{match} is a strict R-matrix and*

$$S^{\mathrm{match}}_{ij} = n - |i - j|$$

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

This means that, given all pairwise pairwise comparions, spectral clustering on S^{match} will recover the true ranking.

Robustness

[Fogel et al., 2014]

Robustness to corrupted entries.

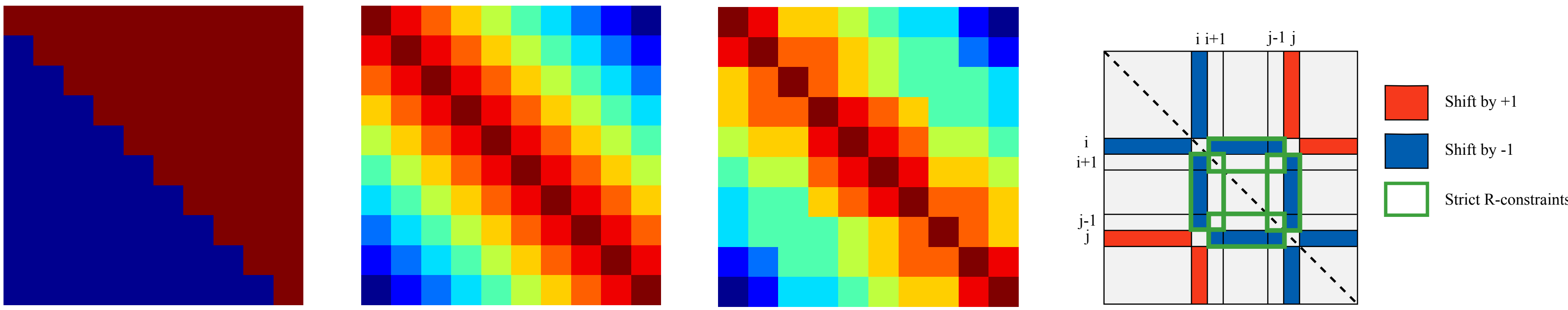
- *Given all comparisons* $C_{s,t} \in \{-1, 1\}$ *between items ordered* $1, \ldots, n$.
- *Suppose the sign of one comparison* $C_{i,j}$ *is switched, with* $i < j$.

If $j - i > 2$ *then* S^{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. . .



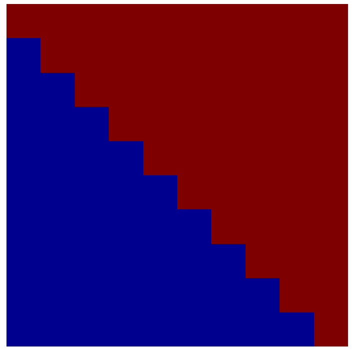
The matrix of pairwise comparisons C *(far left)*.

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

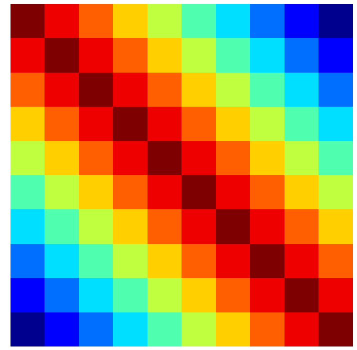
The same S^{match} similarity matrix with comparison (3,8) corrupted *(center right)*. With one corrupted comparison, S^{match} keeps enough strict R-constraints to recover the right permutation. *(far right)*.

Robustness

We can go bit further. . . .



C



S^{match}

- Form S^{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 $x, y \in [0,1]$.

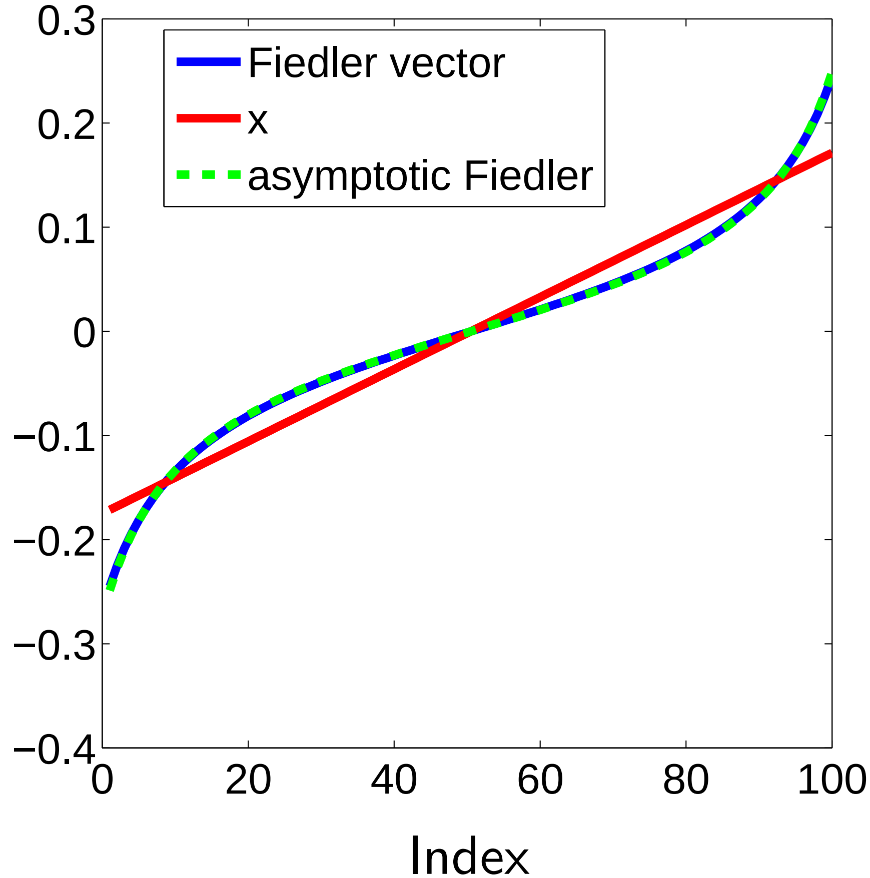
- The degree function is then $d(x) = \int_0^1 k(x,y)dy = -x^2 + x + 1/2$. The range of $d(x)$ is $[0.5, 0.75]$ and the bulk of the spectrum is contained in this interval.
- We can also show that the second smallest eigenvalues of the unnormalized Laplacian satisfies $\lambda_2 < 2/5$, which is outside of this range.
- The **Fiedler vector** f with eigenvalue λ satisfies

$$f''(x)(1/2 - \lambda + x - x^2) + 2f'(x)(1-2x) = 0.$$

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



Comparing the asymptotic Fiedler vector, and the true one for $n = 100$.

Sample optimality

[Fogel et al., 2014]

Sample optimality. *Suppose we observe* $n^{\frac{1}{2}}\mu\nu n \log^3 n$ *consistent comparisons, then*

$$\|\tilde{\pi} - \pi\|_\infty \leq cn/\mu,$$

with probability greater than $1 - \frac{2}{n^{\nu/4-1}}$, *where* μ *controls precision and* $c > 0$ *is an absolute constant.*

Work in progress. . .

Coupon collector: we need at least $O(n \log n)$.

Sample optimality

Proof sketch.

- Bound the Laplacian of the residual $||L_R|| \leq ||D_R|| + ||R||$ where D_R is the degree matrix of R.
- Bound $||R||$ and $||D_R||$ by a function of n and q, with $||L_R|| \leq h(n, q)$.
- Show $|\tilde{\lambda}_3 - \lambda_2| > |\lambda_3 - \lambda_2|/2$ and $|\tilde{\lambda}_1 - \lambda_2| > |\lambda_1 - \lambda_2|/2$ so matrix perturbation results yield

$$||f - \tilde{f}||_2 \leq \sqrt{2}\frac{||\tilde{L} - L||_2}{\min(\lambda_2 - \lambda_1, \lambda_3 - \lambda_2)} = O(||L_R||_2/n^2).$$

- Show eigenvector bound $\|\tilde{f} - f\|_\infty = O(\|\tilde{f} - f\|_2 \sqrt{\log n/n})$.

Sample optimality

- This translates into a bound on the ranking

$$\|\tilde{\pi} - \pi\|_\infty = O(\|\tilde{f} - f\|_\infty n^{3/2}).$$

- Put all pieces together to deduce that

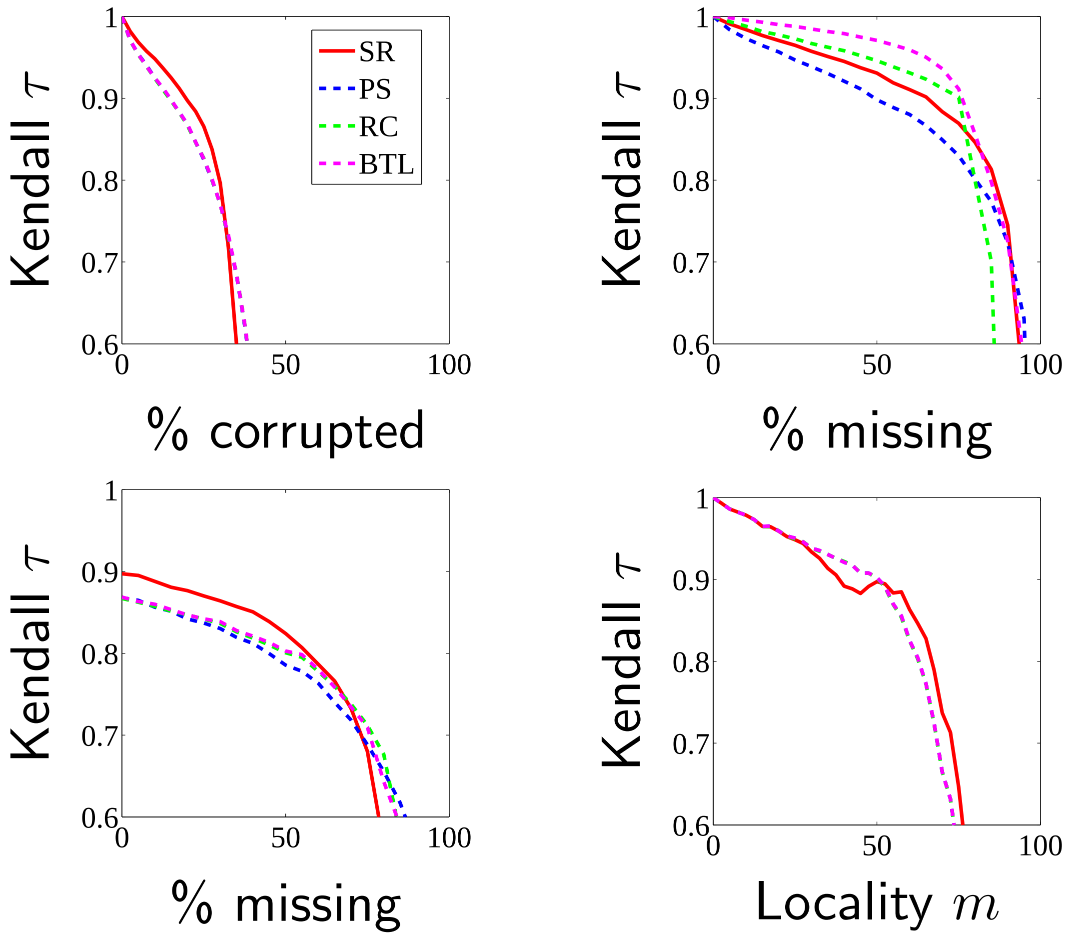
$$\begin{aligned}
\|\tilde{\pi} - \pi\|_\infty &= O(\|\tilde{f} - f\|_\infty n^{3/2}) \\
&= O(\|\tilde{f} - f\|_2 \sqrt{\frac{\log n}{n}} n^{3/2}) \\
&= O(||L_R|| n^{-2} n \sqrt{\log n}) \\
&= O\left(\frac{h(n,q)\sqrt{\log n}}{n}\right).
\end{aligned}$$

When the sampling probability $q \geq \mu\nu\, n^{\frac{1}{2}} n \log^3 n$, we can show $h(n,q) \leq n^2/\sqrt{\log n}$ with high probability.

Outline

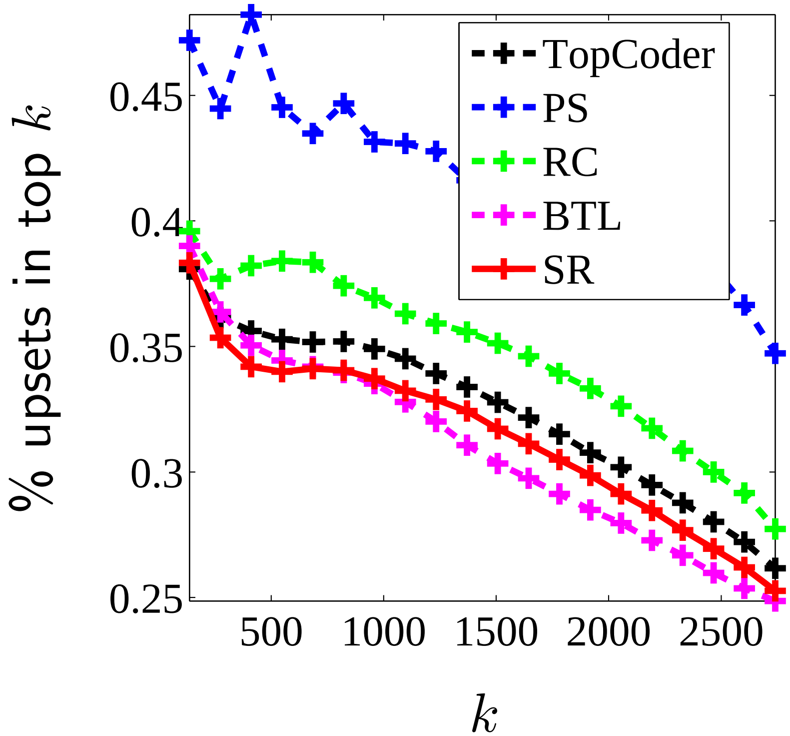
- Introduction
- Seriation and 2-SUM
- Ranking from pairwise comparisons
- **Numerical experiments**

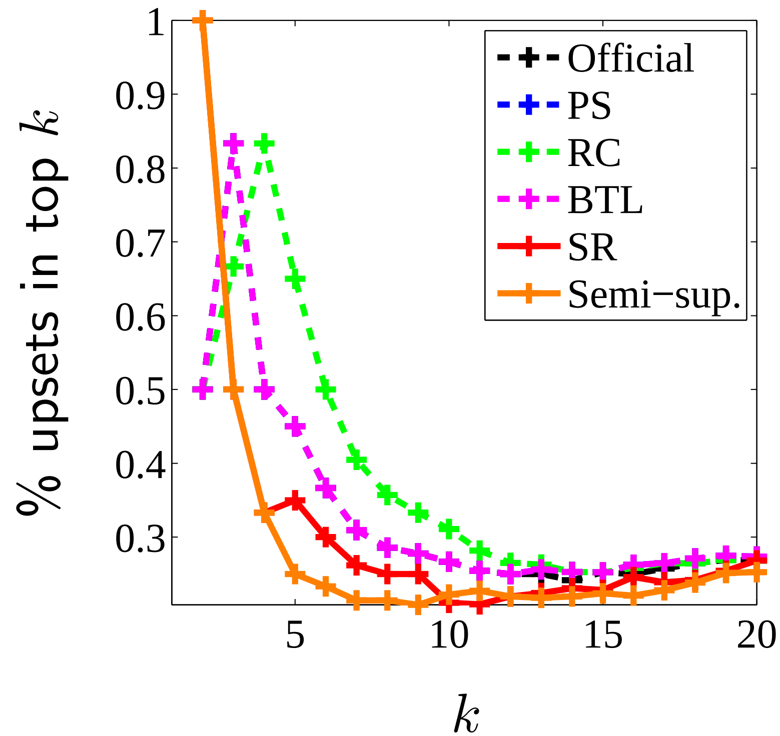
Numerical results



Uniform noise/corruption. Kendall τ (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).

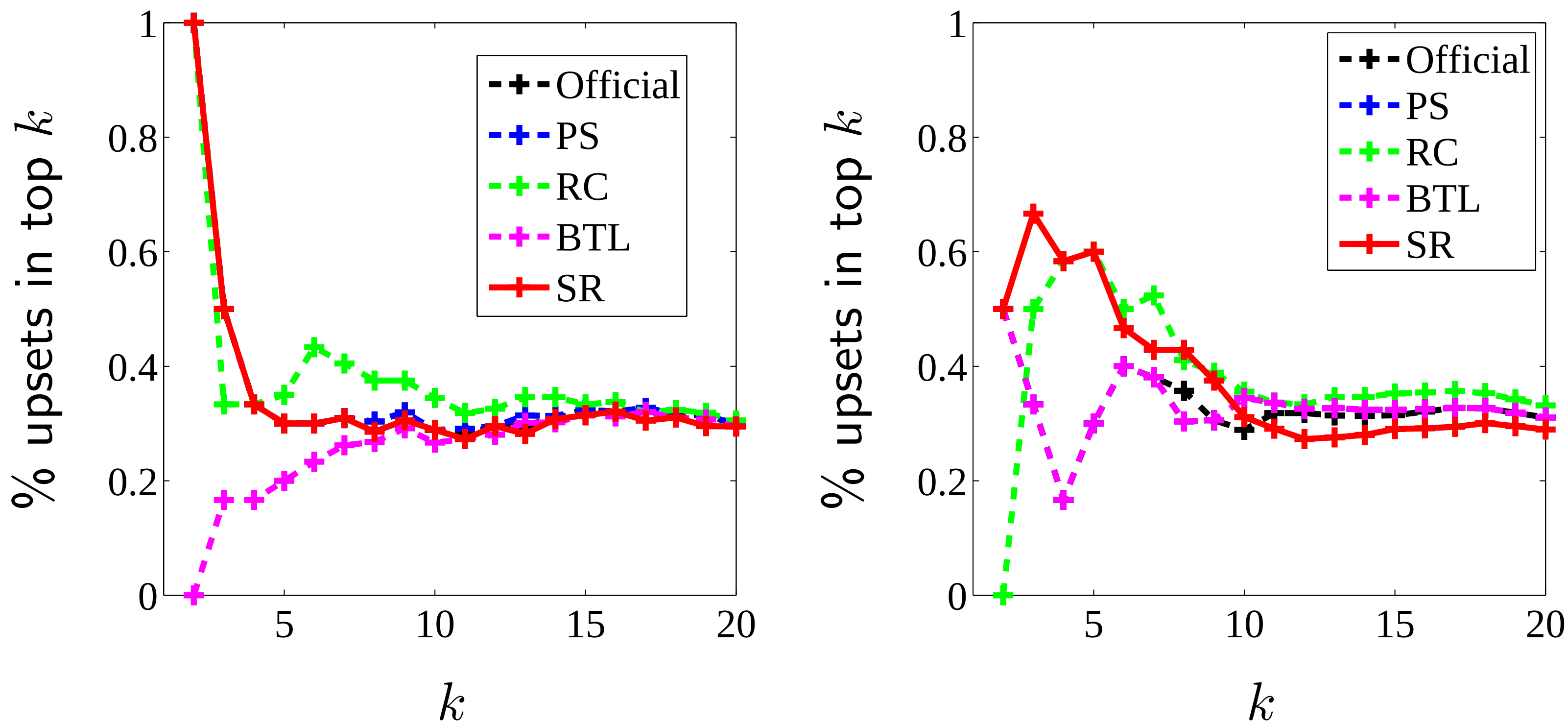
Numerical results





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*).

Numerical results



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*).

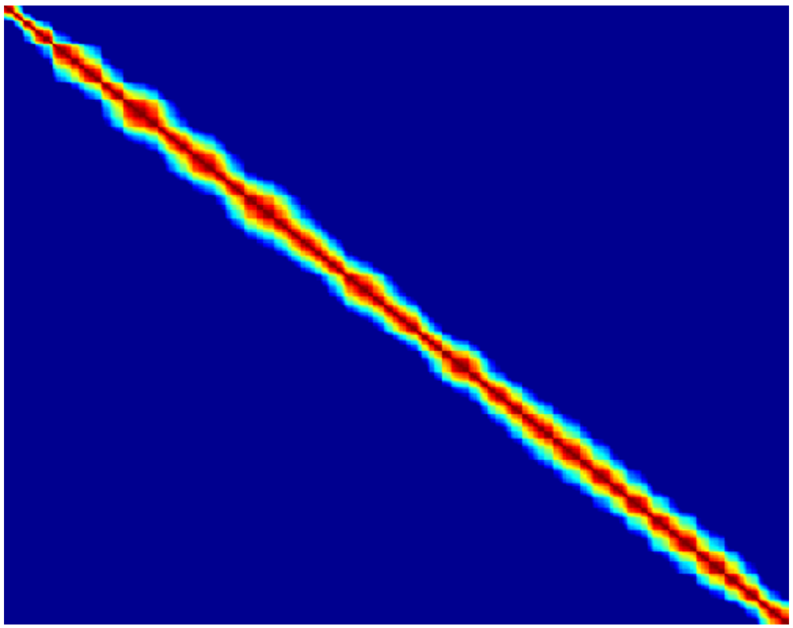
Football teams

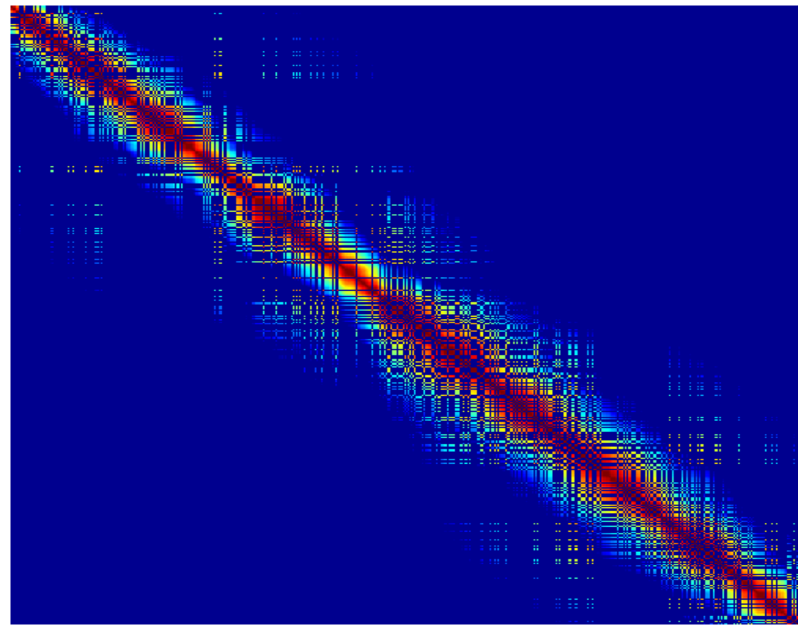
Official	Row-sum	RC	BTL	SerialRank	Semi-Supervised
Man City (86)	Man City	Liverpool	Man City	Man City	Man City
Liverpool (84)	Liverpool	Arsenal	Liverpool	Chelsea	Chelsea
Chelsea (82)	Chelsea	Man City	Chelsea	Liverpool	Liverpool
Arsenal (79)	Arsenal	Chelsea	Arsenal	Arsenal	Everton
Everton (72)	Everton	Everton	Everton	Everton	Arsenal
Tottenham (69)	Tottenham	Tottenham	Tottenham	Tottenham	Tottenham
Man United (64)	Man United	Man United	Man United	Southampton	Man United
Southampton (56)	Southampton	Southampton	Southampton	Man United	Southampton
Stoke (50)	Stoke	Stoke	Stoke	Stoke	Newcastle
Newcastle (49)	Newcastle	Newcastle	Newcastle	Swansea	Stoke
Crystal Palace (45)	Crystal Palace	Swansea	Crystal Palace	Newcastle	West Brom
Swansea (42)	Swansea	Crystal Palace	Swansea	West Brom	Swansea
West Ham (40)	West Brom	West Ham	West Brom	Hull	Crystal Palace
Aston Villa (38)	West Ham	Hull	West Ham	West Ham	Hull
Sunderland (38)	Aston Villa	Aston Villa	Aston Villa	Cardiff	West Ham
Hull (37)	Sunderland	West Brom	Sunderland	Crystal Palace	Fulham
West Brom (36)	Hull	Sunderland	Hull	Fulham	Norwich
Norwich (33)	Norwich	Fulham	Norwich	Norwich	Sunderland
Fulham (32)	Fulham	Norwich	Fulham	Sunderland	Aston Villa
Cardiff (30)	Cardiff	Cardiff	Cardiff	Aston Villa	Cardiff

Ranking of teams in the England premier league season 2013-2014.

Numerical results

DNA. Reorder the *read* similarity matrix to solve C1P on 250 000 reads from human chromosome 22.



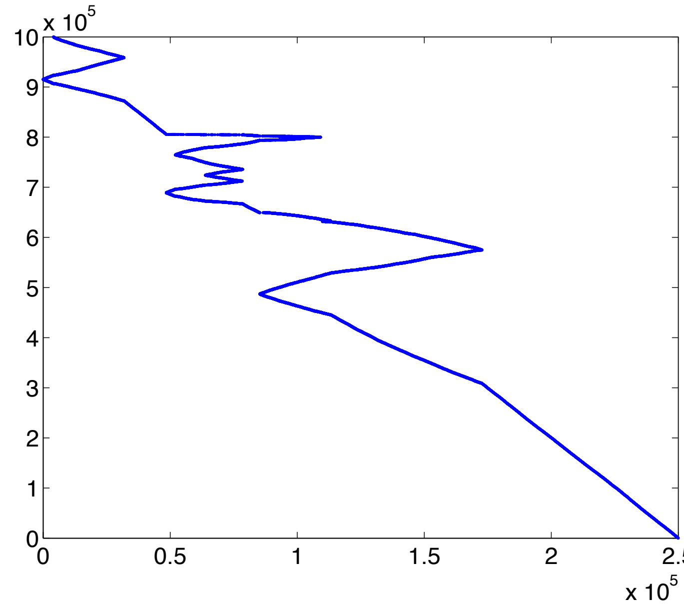


$\# \; reads \times \# \; reads$ matrix measuring the number of common k-mers between read pairs, reordered according to the spectral ordering.

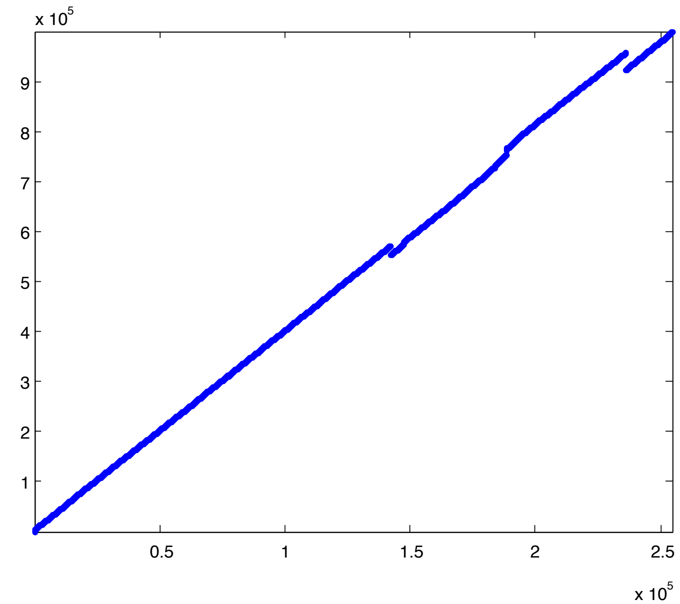
The matrix is $250\,000 \times 250\,000$, we zoom in on two regions.

Numerical results

DNA. 250 000 reads from **human chromosome 22.**



Spectral



Spectral + QP

Recovered read position versus true read position for the **spectral solution** and the **spectral solution followed by semi-supervised seriation**.

We see that the number of misplaced reads significantly decreases in the semi-supervised seriation solution.

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.

NIPS 2014, ArXiv update very soon. . .

*

References

Nir Ailon. Active learning ranking from pairwise preferences with almost optimal query complexity. In *NIPS*, pages 810–818, 2011.

A. Ammar and D. Shah. Ranking: Compare, don't score. *49th Annual Allerton Conference Communication, Control, and Computing*, pages 776–783, 2011.

J.E. Atkins, E.G. Boman, B. Hendrickson, et al. A spectral algorithm for seriation and the consecutive ones problem. *SIAM J. Comput.*, 28 (1):297–310, 1998.

Avrim Blum, Goran Konjevod, R Ravi, and Santosh Vempala. Semidefinite relaxations for minimum bandwidth and other vertex ordering problems. *Theoretical Computer Science*, 235(1):25–42, 2000.

Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, pages 324–345, 1952.

Moses Charikar, Mohammad Taghi Hajiaghayi, Howard Karloff, and Satish Rao. l_2^2 spreading metrics for vertex ordering problems. *Algorithmica*, 56(4):577–604, 2010.

R. Coifman, Y. Shkolnisky, F.J. Sigworth, and A. Singer. Cryo-EM structure determination through eigenvectors of sparse matrices. *working paper*, 2008.

C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. *Proceedings of the Tenth International World Wide Web Conference*, 2001.

Guy Even, Joseph Seffi Naor, Satish Rao, and Baruch Schieber. Divide-and-conquer approximation algorithms via spreading metrics. *Journal of the ACM (JACM)*, 47(4):585–616, 2000.

Uriel Feige. Approximating the bandwidth via volume respecting embeddings. *Journal of Computer and System Sciences*, 60(3):510–539, 2000.

Uriel Feige and James R Lee. An improved approximation ratio for the minimum linear arrangement problem. *Information Processing Letters*, 101(1):26–29, 2007.

F. Fogel, R. Jenatton, F. Bach, and A. d'Aspremont. Convex relaxations for permutation problems. *NIPS 2013, arXiv:1306.4805*, 2013.

F. Fogel, A. d'Aspremont, and M. Vojnovic. Spectral ranking using seriation. *NIPS 2014*, 2014.

Michel X. Goemans. Smallest compact formulation for the permutahedron. *Mathematical Programming*, pages 1–7, 2014.

Ralf Herbrich, Tom Minka, and Thore Graepel. Trueskill$^{\mathrm{TM}}$: A bayesian skill rating system. In *Advances in Neural Information Processing Systems*, pages 569–576, 2006.

Kevin G Jamieson and Robert D Nowak. Active ranking using pairwise comparisons. In *NIPS*, volume 24, pages 2240–2248, 2011.

David G Kendall. Incidence matrices, interval graphs and seriation in archeology. *Pacific Journal of mathematics*, 28(3):565–570, 1969.

Claire Kenyon-Mathieu and Warren Schudy. How to rank with few errors. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 95–103. ACM, 2007.

J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46:604–632, 1999.

Monique Laurent and Matteo Seminaroti. The quadratic assignment problem is easy for robinsonian matrices. *arXiv preprint arXiv:1407.2801*, 2014.

Cong Han Lim and Stephen J Wright. Beyond the birkhoff polytope: Convex relaxations for vector permutation problems. *arXiv preprint arXiv:1407.6609*, 2014.

RD Luce. *Individual choice behavior*. Wiley, 1959.

Sahand Negahban, Sewoong Oh, and Devavrat Shah. Iterative ranking from pairwise comparisons. In *NIPS*, pages 2483–2491, 2012.

A. Nemirovski. Sums of random symmetric matrices and quadratic optimization under orthogonality constraints. *Mathematical programming*, 109(2):283–317, 2007.

L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *Stanford CS Technical Report*, 1998.

Satish Rao and Andréa W Richa. New approximation techniques for some linear ordering problems. *SIAM Journal on Computing*, 34(2): 388–404, 2005.

Thomas L Saaty. A scaling method for priorities in hierarchical structures. *Journal of mathematical psychology*, 15(3):234–281, 1977.

Anthony Man-Cho So. Moment inequalities for sums of random matrices and their applications in optimization. *Mathematical programming*, 130(1):125–151, 2011.

Ulrike Von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pages 555–586, 2008.

Fabian L Wauthier, Michael I Jordan, and Nebojsa Jojic. Efficient ranking from pairwise comparisons. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013.