Ranking & Seriation: a spectral approach

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

with Fajwel Fogel, Rodolphe Jenatton & Francis Bach (INRIA & ENS Paris),
Milan Vojnovic, (MSR Cambridge).

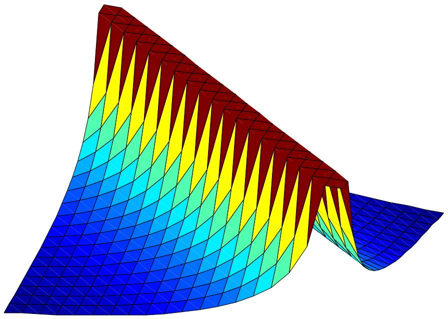
Seriation

The Seriation Problem.

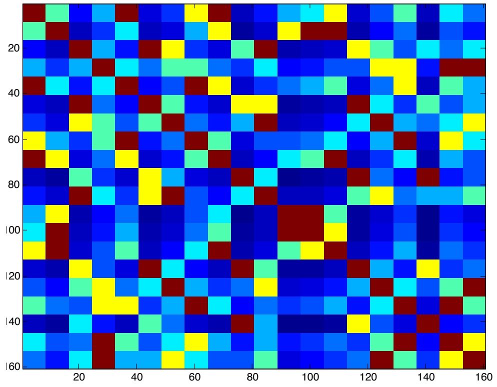
- Pairwise **similarity information** A_{ij} on n variables.
- Suppose the data has a **serial structure**, i.e. there is an order π such that

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

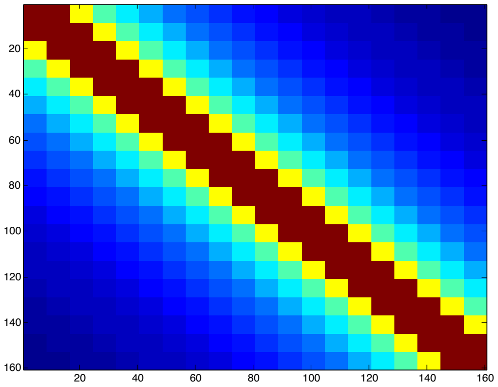
Recover π?



Similarity matrix



Input

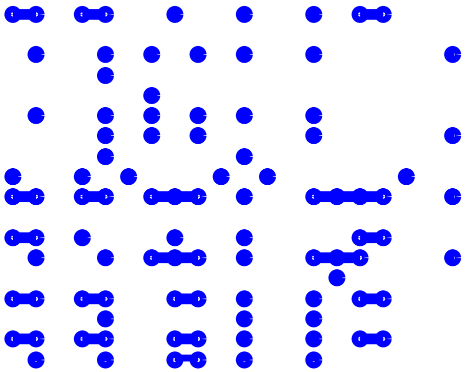


Reconstructed

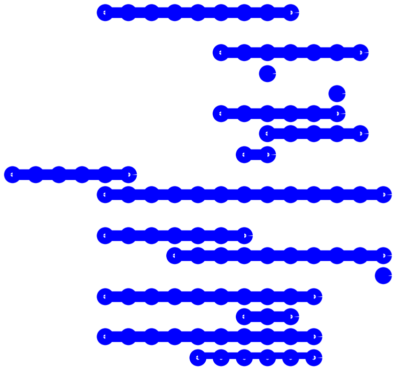
Seriation

The Continuous Ones Problem.

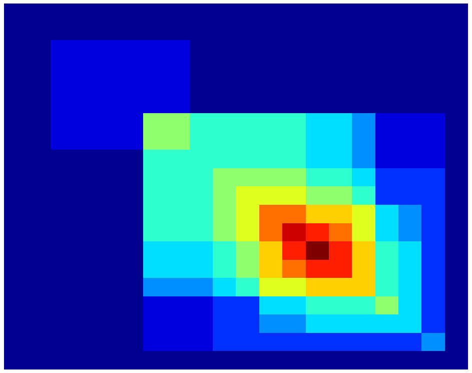
- We're given a rectangular binary $\{0,1\}$ matrix.
- Can we reorder its columns so that the ones in each row are contiguous (C1P)?



Input matrix



Ordered C1P matrix



C^TC (overlap)

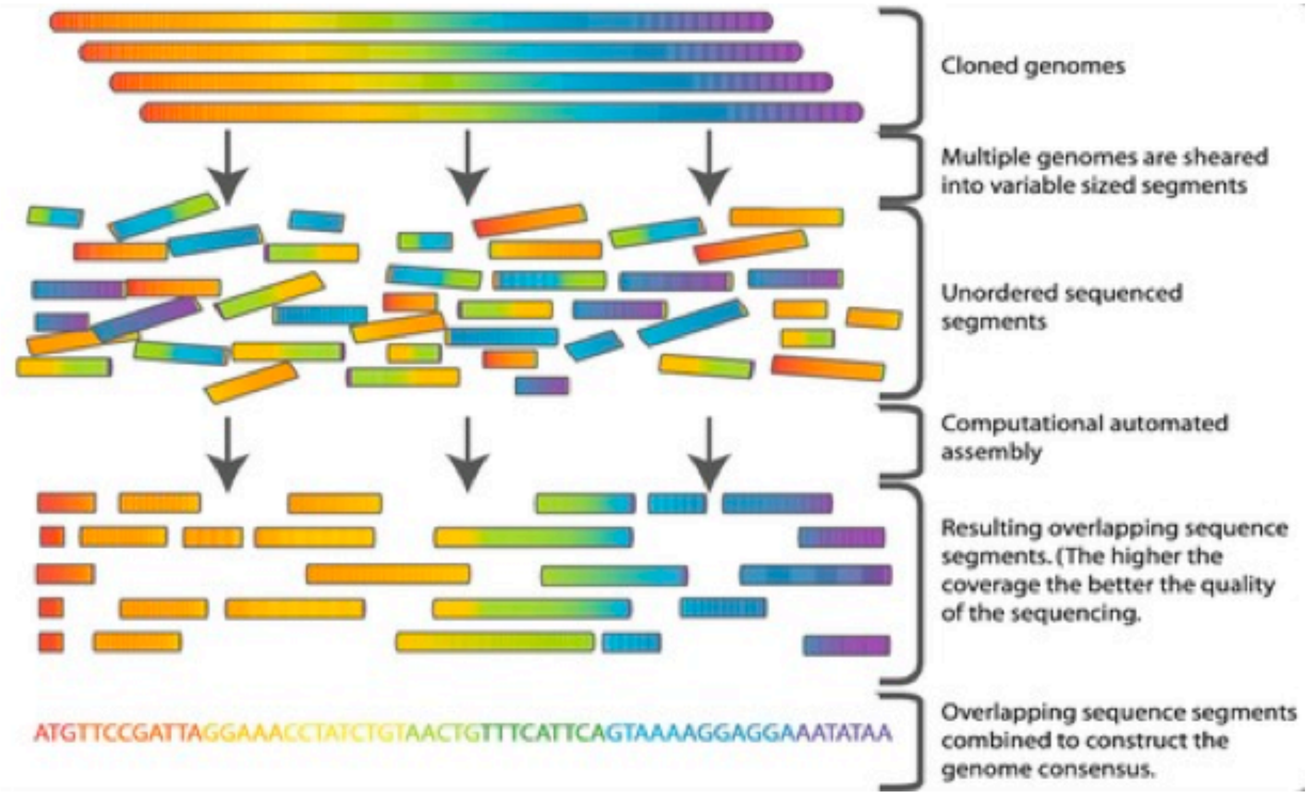
Lemma [Kendall, 1969]

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

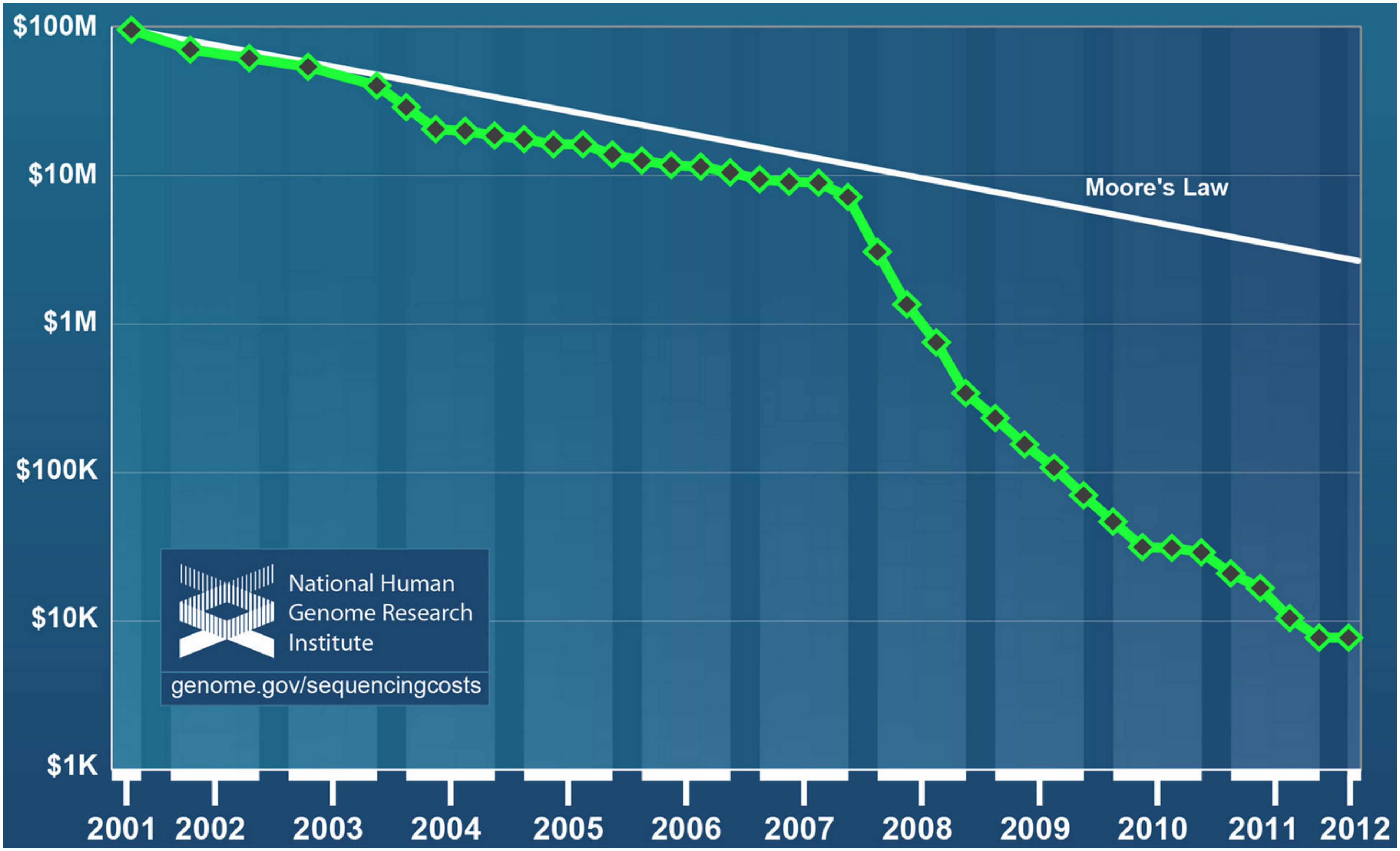
C1P has direct applications in shotgun gene sequencing.

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


Cloned genomes
Multiple genomes are sheared into variable sized segments
Unordered sequenced segments
Computational automated assembly
Resulting overlapping sequence segments. (The higher the coverage the better the quality of the sequencing.
ATGTTCCGATTAGGAAACCTATCTGTAACTGTTTCATTCAGTAAAAGGAGGAAATATAA
Overlapping sequence segments combined to construct the genome consensus.

(from Wikipedia. . .)

Gene Sequencing costs



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

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

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 $A \in \mathbf{S}_n$ 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

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

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 $A = C^2$.*

Convex Relaxation

What's the point?

- Write seriation as an optimization problem.
- Also gives a spectral (hence polynomial) solution for 2-SUM on some R-matrices ([Atkins et al., 1998] mention both problems, but don't show the connection).
- Write a **convex relaxation** for 2-SUM and seriation.
  - Spectral solution scales very well (cf. Pagerank, spectral clustering, etc.)
  - Not very robust. . .
  - Not flexible. . . Hard to include additional structural constraints.

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.**

Convex Relaxation

We 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 $g = (1,\ldots,n)^T$.

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, projecting on doubly stochastic matrices can be done very efficiently, using block coordinate descent on the dual. We use accelerated first-order methods.

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. . .

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

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

- $S^{\text{match}}_{i,j}$ 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*

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

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.

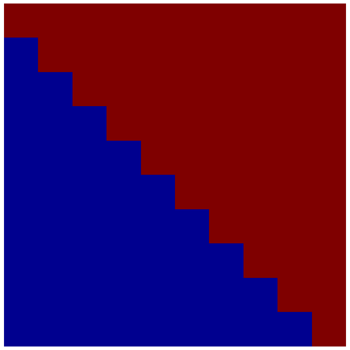
- *Given all comparisons* $C_{s,t} \in \{-1, 1\}$ *between items ordered* $1, \ldots, n$*.*
- *Suppose the sign of one comparison* $C_{i,j}$ *is switched, with* $i < j$*.*

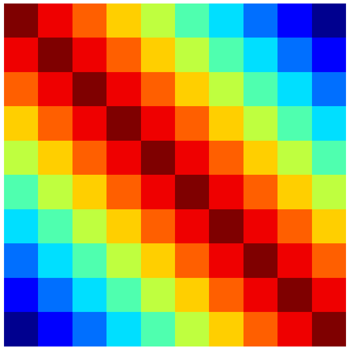
If $j - i > 2$ *then* S^{match} *remains a* strict-R *matrix.*

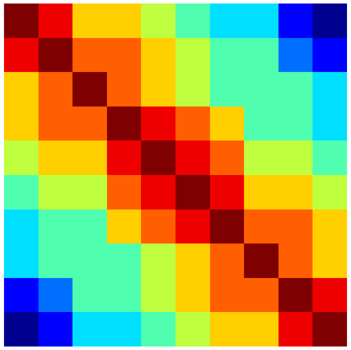
In this case, the score vector w has ties between items i and $i + 1$ and items j and $j - 1$.

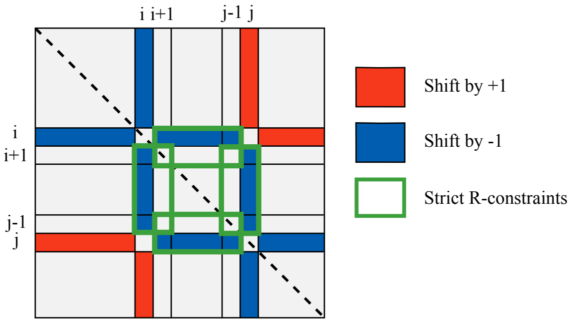
Robustness

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

Generalizes to several errors. . .

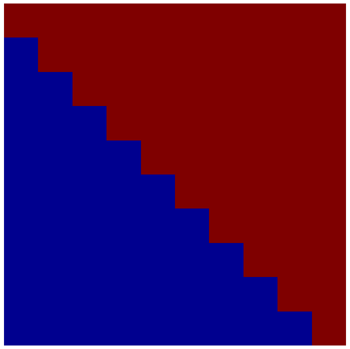
[Fogel et al., 2014]

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^{match} satisfies $p(n, m) \geq 1 - \delta$, provided that $m = O(\sqrt{\delta n})$.*

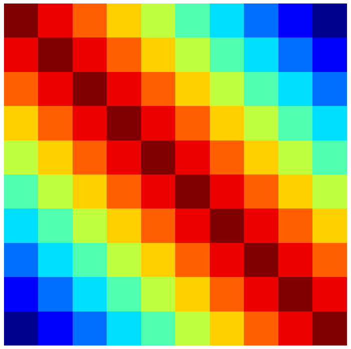
- One corrupted comparison is enough to create ambiguity in scoring arguments.
- Need $\Omega(n^2)$ comparisons for exact recovery [Jamieson and Nowak, 2011].
- No exact recovery results for Markov Chain type spectral methods.

Robustness

We can go bit further. . . .



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

- **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

$$\hat{f} = f - R_2 E f + o(\|E\|_2), \quad \text{with } E = (L-\hat{L})$$

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 $M \in \mathbf{S}_n$ whose entries M are independently sampled as

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

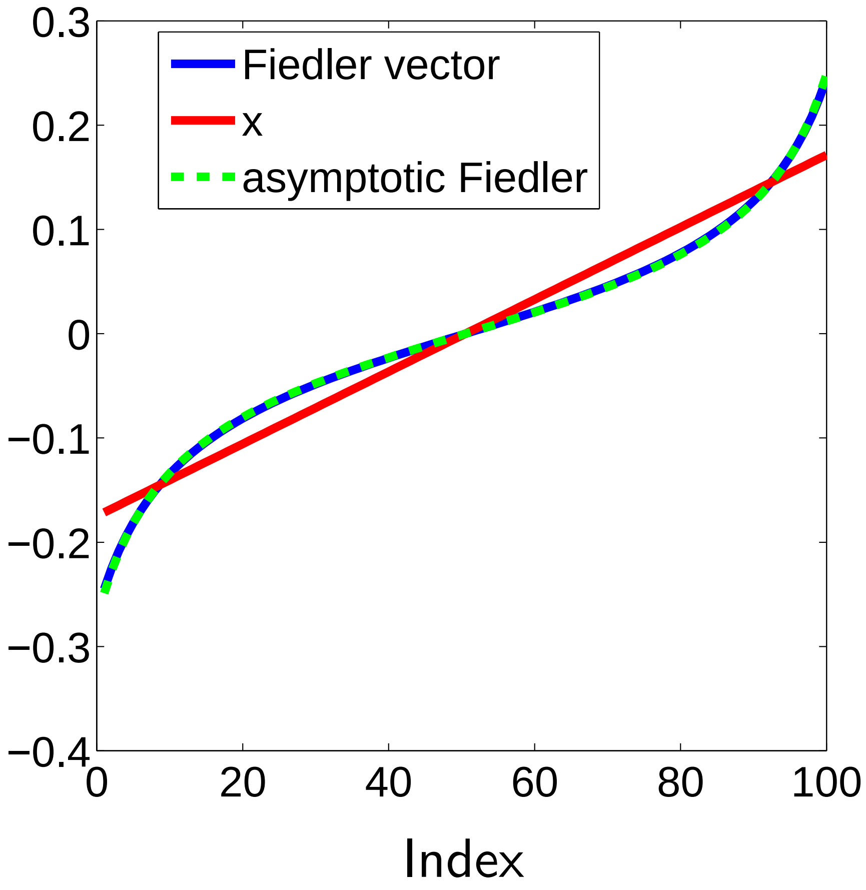
  where $p \in [0,1]$.

- 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.

Robustness

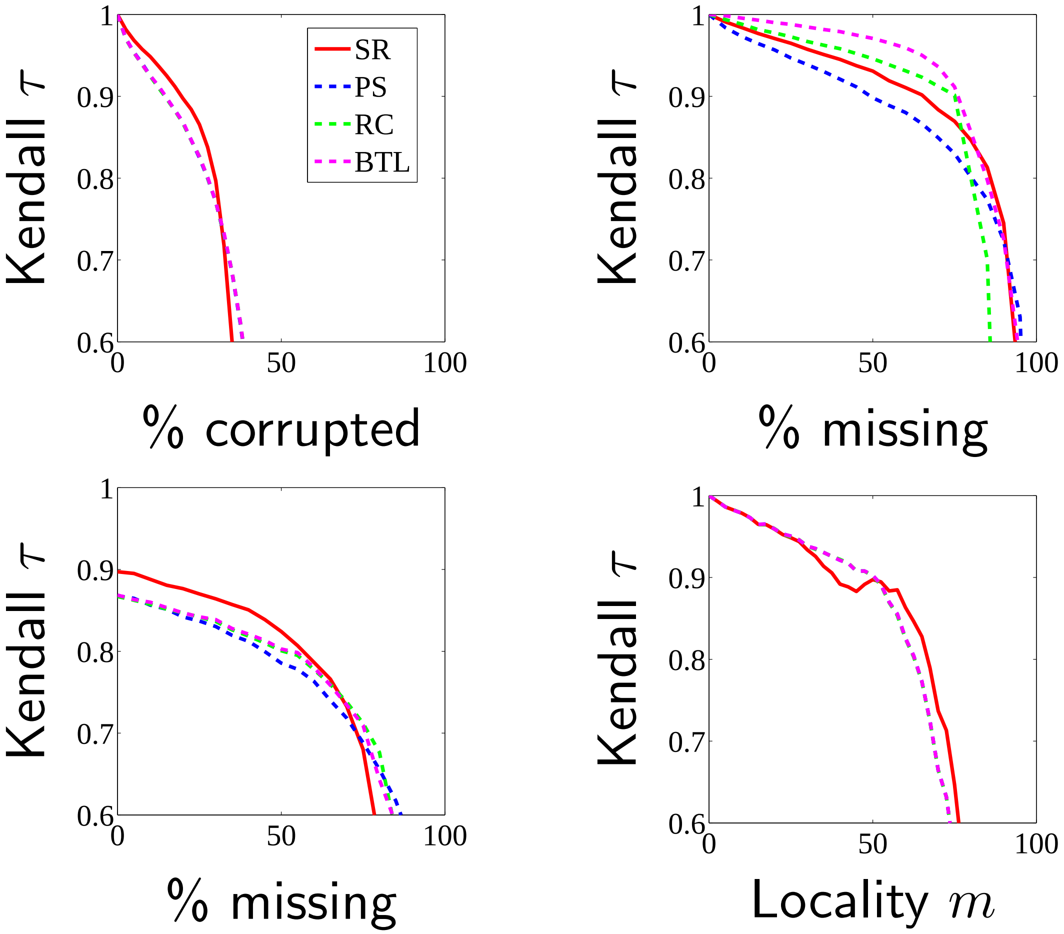


Comparing the asymptotic Fiedler vector, and the true one for $n = 100$.

Outline

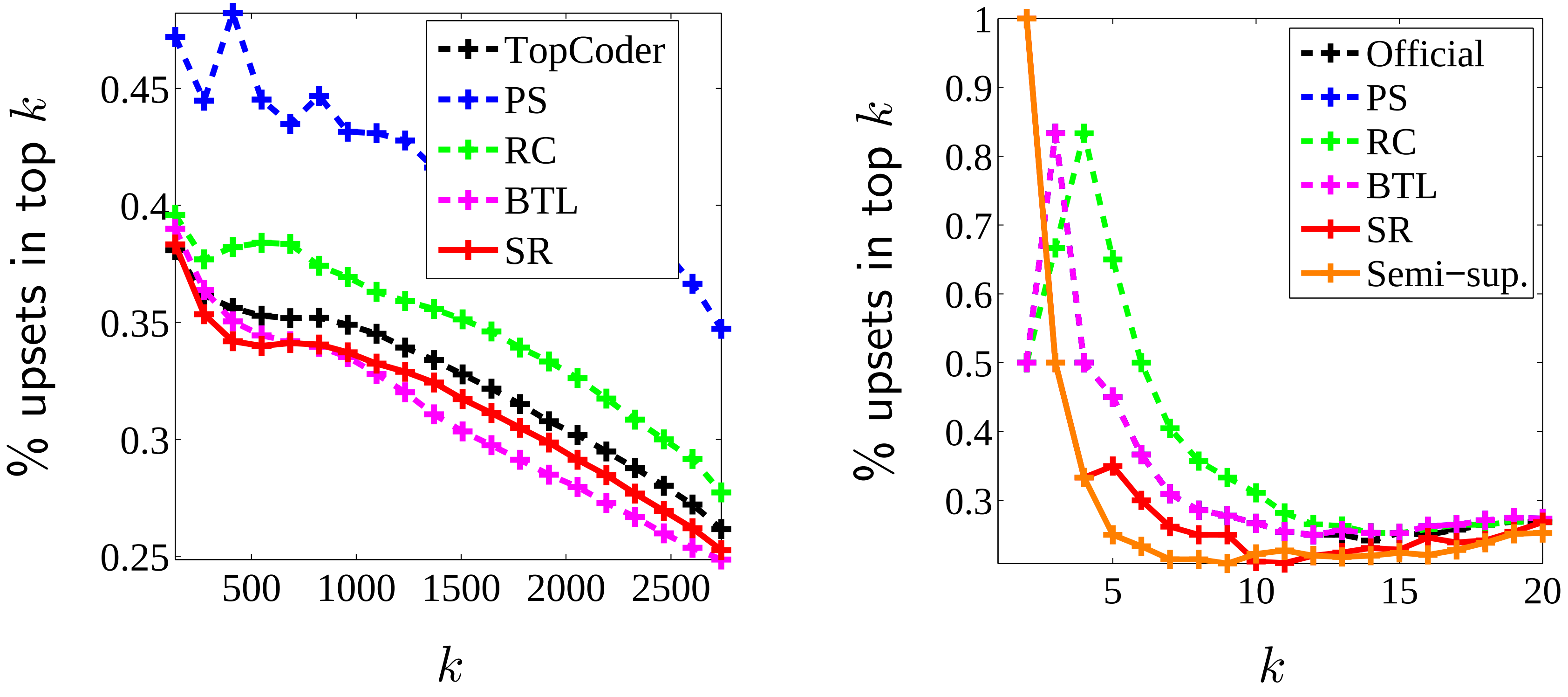
- Introduction
- Seriation and 2-SUM
- Ranking from pairwise comparisons
- **Numerical experiments**

Numerical results



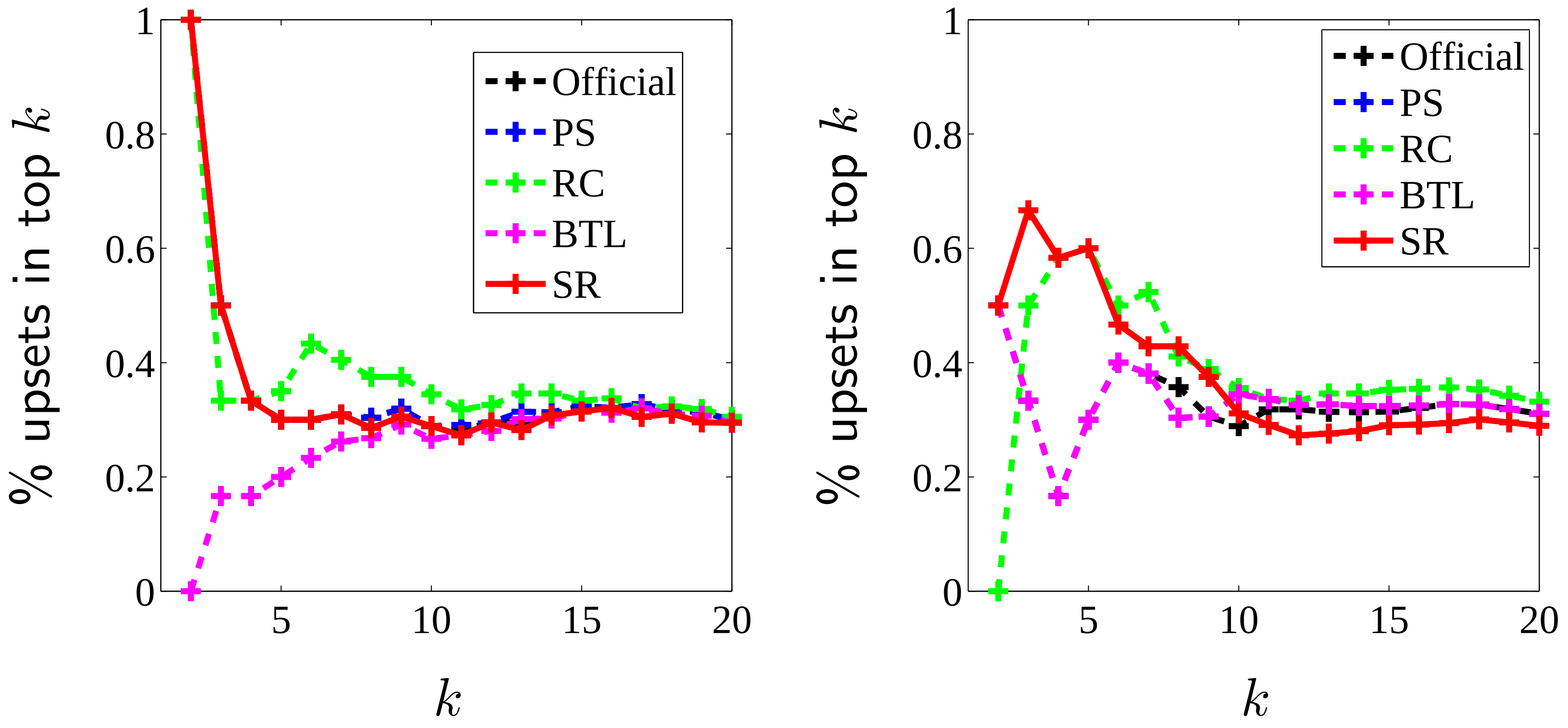
Uniform noise/corruption. Kendall τ (higher is better) for **SerialRank** (SR, full red line), **row-sum** (PS, [Wauthier et al., 2013] dashed blue line), **rank centrality** (RC [Negahban et al., 2012] dashed green line), and **maximum likelihood** (BTL [Bradley and Terry, 1952], dashed magenta line).

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*).

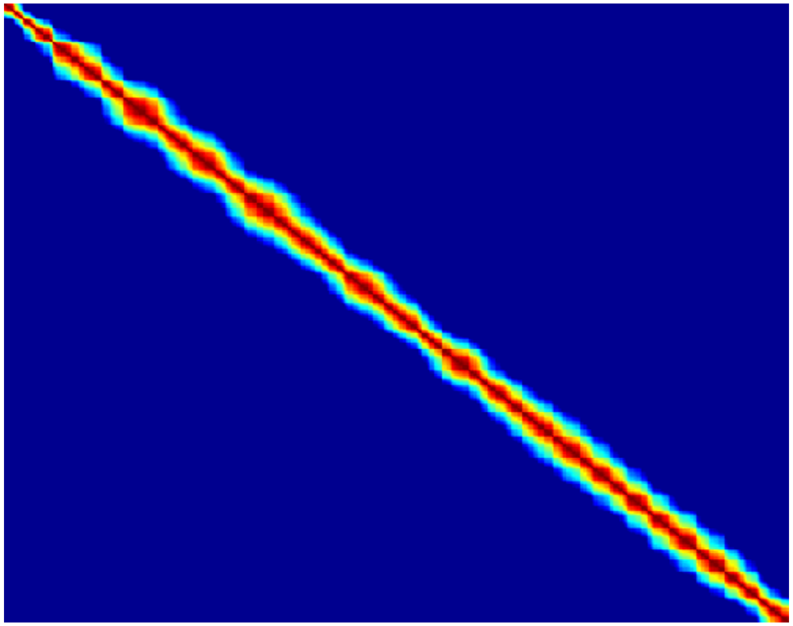
Football teams

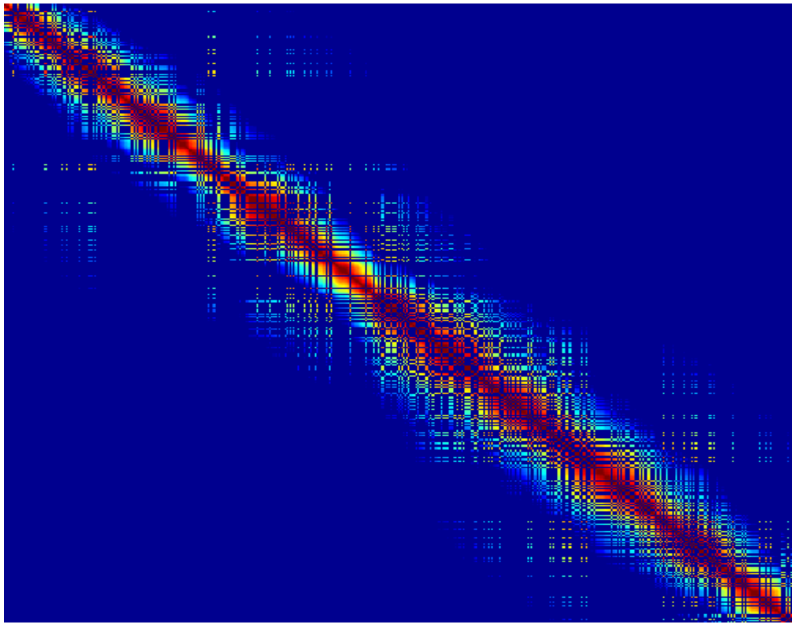
Official	**Row-sum**	**RC**	**BTL**	**SerialRank**	**Semi-Supervised**
Man City (86)	Man City	Liverpool	Man City	Man City	Man City
Liverpool (84)	Liverpool	Arsenal	Liverpool	Chelsea	Chelsea
Chelsea (82)	Chelsea	Man City	Chelsea	Liverpool	Liverpool
Arsenal (79)	Arsenal	Chelsea	Arsenal	Arsenal	Everton
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.



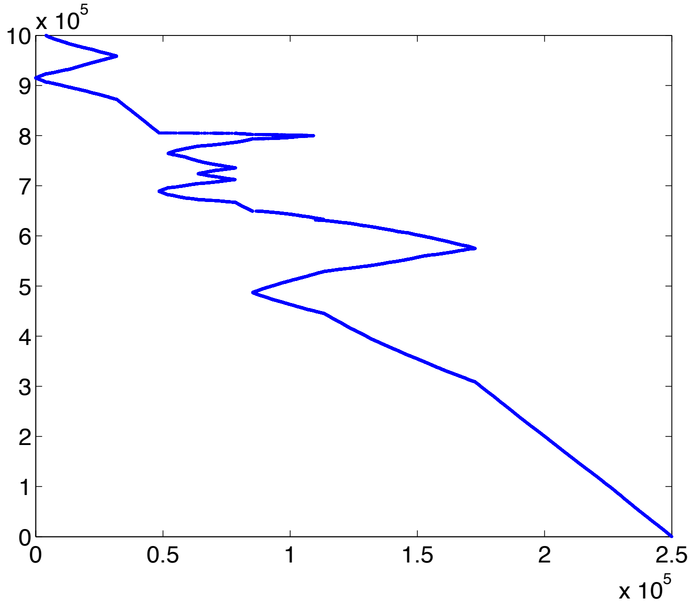


$\# \; reads \times \# \; reads$ matrix measuring the number of common k-mers between read pairs, reordered according to the spectral ordering.

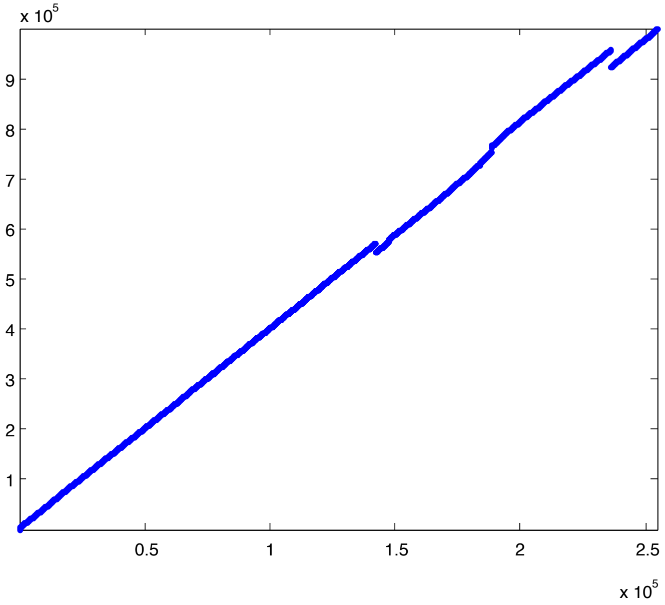
The matrix is 250 000 × 250 000, we zoom in on two regions.

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.

Coming soon. . .

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

NIPS 2014, ArXiv. . .

*

References

D. Achlioptas and F. McSherry. Fast computation of low-rank matrix approximations. *Journal of the ACM*, 54(2), 2007.

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.

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

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.

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.

Ralf Herbrich, Tom Minka, and Thore Graepel. Trueskill$^{\text{TM}}$: 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.

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.

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

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.