Lecture 17: Ranking from pairwise comparisons

Foundations of Data Science: Algorithms and Mathematical Foundations

Mihai Cucuringu
mihai.cucuringu@stats.ox.ac.uk

CDT in Mathematics of Random System
University of Oxford

28 September, 2023

Ranking from pairwise information

n players: incomplete inconsistent pairwise comparisons

$$\text{(ordinal)} \qquad \text{Player}_i \succ \text{Player}_j$$

$$\text{(cardinal)} \qquad \text{Player}_i \;\; 3 \; : \; 1 \;\; \text{Player}_j$$

Goal: infer a global or partial ranking $\pi(i)$ of the n players

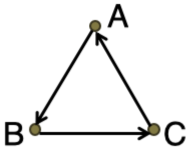
$$\text{Player}_{\pi(1)} \succ \text{Player}_{\pi(2)} \succ \ldots \succ \text{Player}_{\pi(n)}$$

that "best" agrees with the data (eg., minimize the number of upsets)

- $r_1, r_2, \ldots, r_n$ ground truth ranking $r_1, r_2, \ldots, r_n \in \{1, 2, \ldots, n\}$ (or real-valued strengths $r_i \in \mathbb{R}, i = 1, \ldots, n$)
- available pairwise comparisons are a proxy for the rank (or strength) offset $r_i - r_j$
- goal: recover estimates $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_n$ (or at least their relative ordering)

Challenges

- in most practical applications, the available information
  - is usually incomplete, especially when n is large (meaning that we only observe a small subset of pairwise measurements)
  - is very noisy (meaning that most measurements are inconsistent with the existence of an underlying total ordering)
- at sports tournaments ($G = K_n$) the outcomes always contain cycles: A beats B, B beats C, and C beats A



- aim to recover a total (or partial) ordering that is as consistent as possible with the data
- minimize the number of upsets: pairs of players for which the higher ranked player is beaten by the lower ranked one

Challenges

- the available measurements are not uniformly distributed around the network, and can significantly affect the ranking procedure
- the noise in the data may not be distributed uniformly throughout the network, with part of the network containing pairwise measurements that are a lot less noisy than the rest of the network
  - opportunity to recover partial rankings
  - rely on recent spectral algorithms for detecting planted cliques or dense subgraphs in a graph

Why ranking?

- Ranking is a central part of many information retrieval problems.

The analysis of many modern large-scale data sets implicitly requires various forms of ranking to allow for

- the identification of the most important entries
- efficient computation of search and sort operations
- extraction of main features

Instances of such problems are abundant in various disciplines, especially in modern internet-related applications such as

- the famous search engine provided by Google
- eBay's feedback-based reputation mechanism
- Amazon's Mechanical Turk (MTurk) crowdsourcing to coordinate the use of human labor to perform various tasks
- Netflix movie recommendation system
- Cite-Seer network of citations
- ranking of college football/basketball teams

Exchange economic systems

- A purely exchange economic system may be described by a graph $G = (V, E)$ with vertex set $V = \{1, ..., n\}$ representing the n goods and edge set E representing feasible pairwise transactions
- If the market is complete (every pair of goods is exchangeable), then $G = K_n$
- Exchange rate between the i^{th} and j^{th} goods is

$$1 \text{ unit } i = a_{ij}, \; a_{ij} > 0$$

- Exchange rate matrix A = $[a_{ij}]$ is a reciprocal matrix (possibly incomplete)
- Used for paired preference aggregation, and later on, for currency exchange analysis
- The problem of pricing is to look for a universal equivalent that measures the values of goods $\pi : V \mapsto \mathbb{R}$ s.t.

$$a_{ij} = \frac{\pi_j}{\pi_i}$$

Exchange economic systems

- The problem of pricing is to look for a universal equivalent that measures the values of goods $\pi : V \mapsto \mathbb{R}$ s.t.

$$a_{ij} = \frac{\pi_j}{\pi_i}$$

- In complete markets, there exists a universal equivalent if and only if the market is triangular arbitrage-free $a_{ij}a_{jk} = a_{ik}$, for all distinct $i, j, k \in V$
- Transform into a pairwise ranking problem via the logarithmic map

$$X_{ij} = \log a_{ij}$$

- C_3 arbitrage-free equivalent to $X_{ij} + X_{jk} + X_{ki} = 0$
- Thus, asking if a universal equivalent exists is equivalent to asking if a global ranking $s : V \mapsto \mathbb{R}$ exists so that

$$X_{ij} = s_j - s_i$$

with $s_i = \log \pi_i$

Very rich literature on ranking

- dates back as early as the 1940s (Kendall and Smith)
- **PageRank**: used by Google to rank web pages in increasing order of their relevance (see previous lecture)
- Kleinberg's **HITS algorithm**: another website ranking algorithm based on identifying good *authorities* and *hubs* for a given topic queried by the user

Traditional ranking methods fall short:

- developed with **ordinal** comparisons in mind (movie X is better than movie Y)
- much of the current data deals with **cardinal/numerical** scores for the pairwise comparisons (e.g., goal difference in sports)

Erdős-Rényi Outliers noise model

$r_1, \ldots, r_n$ denote the ground truth rankings of the n players

ERO(n, p, η): the available measurements are given by

$$C_{ij} = \begin{cases} r_i - r_j & \text{correct edge} & \text{w.p. } (1-\eta)p \\ \sim \text{Unif}[-(n-1), n-1] & \text{incorrect edge} & \text{w.p. } \eta p \\ 0 & \text{missing edge,} & \text{w.p. } 1-p \end{cases} \tag{1}$$

Multiplicative Uniform Noise model

MUN(n, p, η): noise is multiplicative and uniform

- for cardinal measurements, instead of the true rank-offset measurement $r_i - r_j$, we measure

$$C_{ij} = (r_i - r_j)(1 + \epsilon), \quad \text{where} \quad \epsilon \sim [-\eta, \eta]. \tag{2}$$

- cap the erroneous measurements at $n - 1$ in magnitude
- for ordinal measurements, $C_{ij} = \text{sign}\left((r_i - r_j)(1 + \epsilon)\right)$

E.g., $\eta = 50\%$, and $r_i - r_j = 10$, then $C_{ij} \sim [5, 15]$.

Serial-Rank (NIPS 2014; JMLR 2016)

$$C_{ij} = \begin{cases} 1 & \text{if i is ranked higher than j} \\ 0 & \text{if i and j are tied, or comparison is not available} \\ -1 & \text{if j is ranked higher than i} \end{cases} \tag{3}$$

- the pairwise similarity matrix is given by

$$S_{ij}^{match} = \sum_{k=1}^{n} \left(\frac{1 + C_{ik} C_{jk}}{2} \right) \tag{4}$$

- $C_{ik} C_{jk} = 1$ whenever i and j have the same signs, and $C_{ik} C_{jk} = -1$ whenever they have opposite signs
- S_{ij}^{match} counts the number of matching comparisons between i and j with a third reference item k
- intuition: players that beat the same players and are beaten by the same players should have a similar ranking in the final solution

$$S^{match} = \frac{1}{2} \left(n\mathbf{1}\mathbf{1}^T + CC^T \right) \tag{5}$$

Algorithm 1 Serial-Rank: an algorithm for spectral ranking using seriation, proposed by Fogel, d'Aspremont and Vojnovic

Require: A set of pairwise comparisons $C_{ij} \in \{-1, 0, 1\}$ or [-1,1]

1: Compute a similarity matrix as shown in (4)

2: Compute the associated graph Laplacian matrix

$$L_S = D - S \tag{6}$$

for diagonal matrix $D = \mathbf{diag}\,(S\mathbf{1})$; $D_{ii} = \sum_{j=1}^{n} S_{i,j}$ is the deg(i)

3: Compute the Fiedler vector of S (eigenvector corresponding to the smallest nonzero eigenvalue of L_S).

4: Output the ranking induced by sorting the Fiedler vector of S, with the global ordering (increasing or decreasing order) chosen such that the number of upsets is minimized.

Serial-Rank: Robustness to corrupted entries

Theorem (Fogel et al., 2014)

Given a comparison matrix for n items with m corrupted comparisons selected uniformly at random from the set of all possible item pairs. The prob. of exact recovery $p(n, m)$ using seriation on S^{match} satisfies $p(n, m) \geq 1 - \delta$, if $m = O(\sqrt{\delta n})$.

Rank Centrality (NIPS 2012)

- Negahban, Oh, Shah, "*Rank Centrality: Ranking from Pair-wise Comparisons*", NIPS 2012
- algorithm for rank aggregation by estimating scores for the items from the stationary distribution of a certain random walk on the graph of items
- edges encode the outcome of pairwise comparisons
- proposed for the rank aggregation problem: a collection of sets of pairwise comparisons over n players, given by k different ranking systems

Rank Centrality

- the probability of transitioning from vertex i to vertex j is directly proportional to how often player j beat player i across all the matches played

- the random walk has a higher chance of transitioning to a more skillful neighbors

- the frequency of visiting a particular node, which reflects the rank or the skill level of the corresponding players, is encoded in the stationary distribution of the associated Markov Chain

- $Y_{ij}^{(l)} = 1$ if player j beats player i, and 0 otherwise, during the l^{th} match, $l = 1, \ldots, k$

Rank Centrality; $k > 1$ rating systems

- the famous BTL (Bradley-Terry-Luce) model assumes that

$$\mathbb{P}(Y_{ij}^{(l)} = 1) = \frac{w_j}{w_i + w_j} \tag{7}$$

- w is the vector of positive weights associated to the players
- RC first estimates the fraction of times j defeated i (in k matches)

$$a_{ij} = \frac{1}{k} \sum_{l=1}^{k} Y_{ij}^{(l)}$$

- consider the symmetric matrix

$$A_{ij} = \frac{a_{ij}}{a_{ij} + a_{ji}} \tag{8}$$

$$P_{ij} = \begin{cases} \frac{1}{d_{max}} A_{ij} & \text{if } i \neq j \\ 1 - \frac{1}{d_{max}} \sum_{k \neq i} A_{ik} & \text{if } i = j, \end{cases} \tag{9}$$

  where d_{max} denotes the maximum out-degree of a node.
- recover the scores of the players/items items from the stationary distribution of P (an eigenvector calculation)

Singular Value Decomposition (SVD) ranking

- for cardinal measurements $C_{ij} = r_i - r_j$, the noiseless matrix of rank offsets $C = (C_{ij})_{1 \leq i,j \leq n}$ is a skew-symmetric of even rank 2
$$C = r\boldsymbol{e}^T - \boldsymbol{e}r^T \tag{10}$$
where $\boldsymbol{e}$ denotes the all-ones column vector (check this!)
- under the ERO noise model (also check)
$$\mathbb{E}C_{ij} = (r_i - r_j)(1-\eta)p, \tag{11}$$
- in matrix form: $\mathbb{E}C$ is a rank-2 skew-symmetric matrix
$$\mathbb{E}C = (1-\eta)p(r\boldsymbol{e}^T - \boldsymbol{e}r^T) \tag{12}$$
- can decompose the given data matrix C as
$$C = \mathbb{E}C + R \tag{13}$$
- where R is a random noise matrix
- can recover the ordering from top 2 singular vectors of C: $\{v_1, v_2, -v_1, -v_2\}$; order their entries, infer rankings, and choose whichever minimizes the number of upsets
- amenable to a theoretical analysis using tools from random matrix theory on rank-2 deformations of random matrices.

Ranking via Least-Squares

- $m = |E(G)|$
- recall that instead of clean measurements
$$T_{ij} = r_i - r_j \tag{14}$$
we observe noisy measurements
$$\underbrace{T_{ij} + \text{noise}}_{C_{ij}} = r_i - r_j \tag{15}$$
thus we can set up a linear system of equations
$$r_i - r_j \approx C_{ij} \tag{16}$$
and solve this in the least squares sense.
- denote by B the edge-vertex incidence matrix of size $m \times n$
$$B_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E(G), \quad \text{and } i > j \\ -1 & \text{if } (i,j) \in E(G), \quad \text{and } i < j \\ 0 & \text{if } (i,j) \notin E(G) \end{cases} \tag{17}$$
- w the vector of size $m \times 1$ containing all pairwise comparisons $w(e) = C_{ij}$, for all edges $e = (i,j) \in E(G)$
- least-squares solution to the ranking problem
$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad ||Bx - w||_2^2. \tag{18}$$

From synchronization to ranking

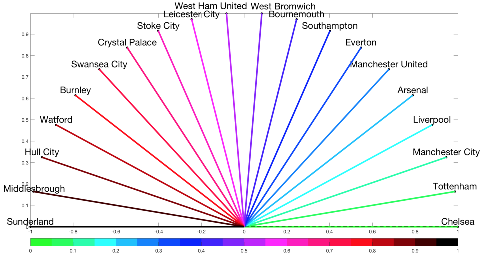
Estimate n unknown angles $\theta_1, \ldots, \theta_n \in [0, 2\pi)$, given m noisy measurements Θ_{ij} of their pairwise offsets

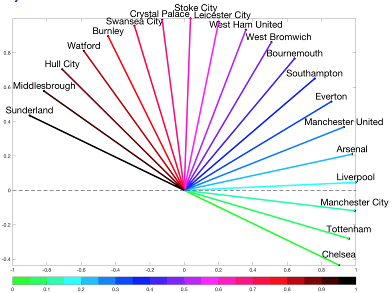
$$\Theta_{ij} = \theta_i - \theta_j \mod 2\pi, \quad (ij) \in E(G) \tag{19}$$

Estimate n unknown strength/ranks $r_1, \ldots, r_2 \in [0, M]$, given m noisy measurements C_{ij} of their pairwise offsets

$$C_{ij} = r_i - r_j, \quad (ij) \in E(G) \tag{20}$$

Synchronization Ranking (Sync-Rank)





Map all rank offsets C_{ij} to an angle $\Theta_{ij} \in [0, 2\pi\delta)$

$$C_{ij} \mapsto \Theta_{ij} := 2\pi\delta \frac{C_{ij}}{n-1} \tag{21}$$

$$H_{ij} = \begin{cases} e^{\imath \Theta_{ij}} & \text{if } (i,j) \in E \\ 0 & \text{if } (i,j) \notin E. \end{cases} \tag{22}$$

$$\underset{\theta_1,\ldots,\theta_n \in [0,2\pi)}{\text{maximize}} \sum_{i,j=1}^{n} e^{-\imath\theta_i} H_{ij} e^{\imath\theta_j} \quad + \quad \text{relax (spectral or SDP)} \tag{23}$$

M. Cucuringu, *Sync-Rank: Robust Ranking, Constrained Ranking and Rank Aggregation via Eigenvector and SDP Synchronization*, IEEE Transactions on Network Science and Engineering (2016)

Comparison of several algorithms

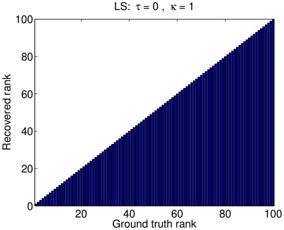
Acronym	Name
SVD	SVD Ranking
LS	Least Squares Ranking
SER	Serial-Ranking (NIPS 2014)
SER-GLM	Serial-Ranking in the GLM model (NIPS 2014)
RC	Rank-Centrality (NIPS 2012)
SYNC	Sync-Rank via the spectral relaxation
SYNC-SDP	Sync-Rank via SDP relaxation

Table: Names of the algorithms we compare, and their acronyms.

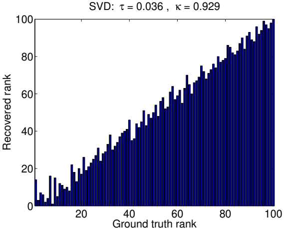
Let us look at recovery error, as we vary

- the noise level $\eta = \{0, 0.35, 0.75\}$ in the measurements
- at a fixed sparsity $p = 0.5$ of the measurement graph.

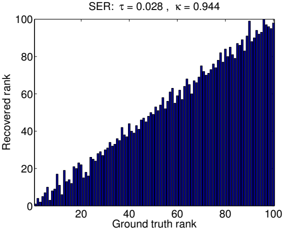
ERO, $\eta = 0$, $p = 0.5$



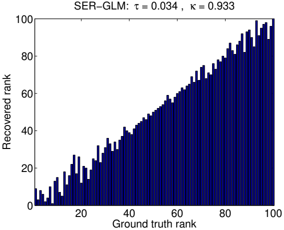
(a) LS



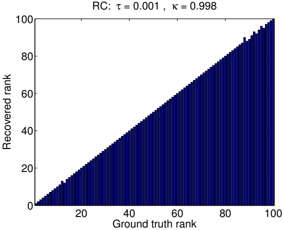
(b) SVD



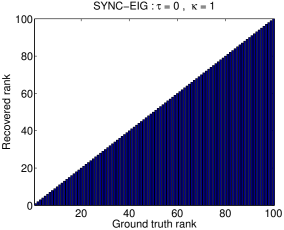
(c) SER



(d) SER-GLM



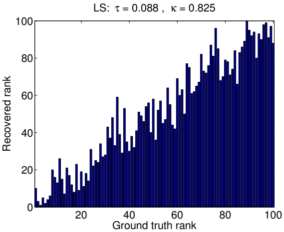
(e) RC



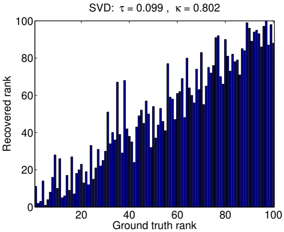
(f) SYNC-EIG

- For ease of visualization, the $n = 100$ player strengths/ranks are such that $r_i = i, i = 1, 2, \quad \ldots, n$.

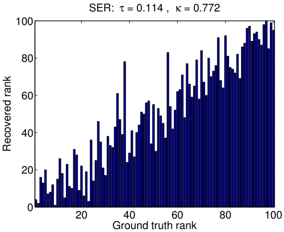
ERO, $\eta = 0.35$, $p = 0.5$



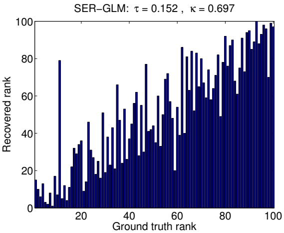
(g) LS



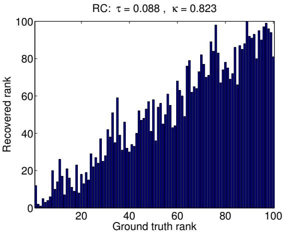
(h) SVD



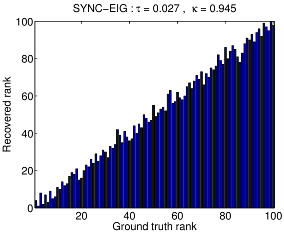
(i) SER



(j) SER-GLM

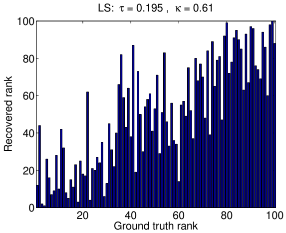


(k) RC

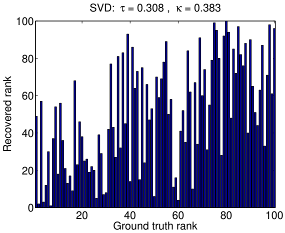


(l) SYNC-EIG

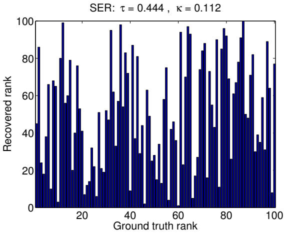
ERO, $\eta = 0.75$, $p = 0.5$



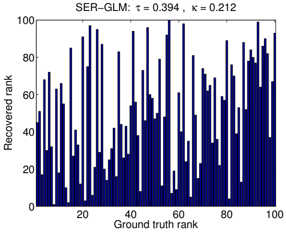
(m) LS



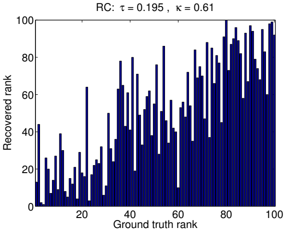
(n) SVD



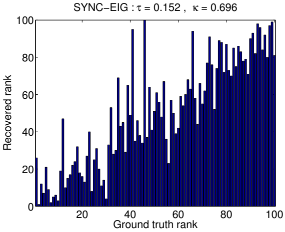
(o) SER



(p) SER-GLM



(q) RC



(r) SYNC-EIG

Kendall distance

- measure accuracy using the popular Kendall distance (Lecture 3)
- counts the number of pairs of candidates that are ranked in different order (flips), in the two permutations (the original one and the recovered one)

$$\kappa(\pi_1, \pi_2) = \frac{|\{(i,j) : i < j, [\pi_1(i) < \pi_1(j) \ \wedge \ \pi_2(i) > \pi_2(j)] \ \}}{\binom{n}{2}}$$

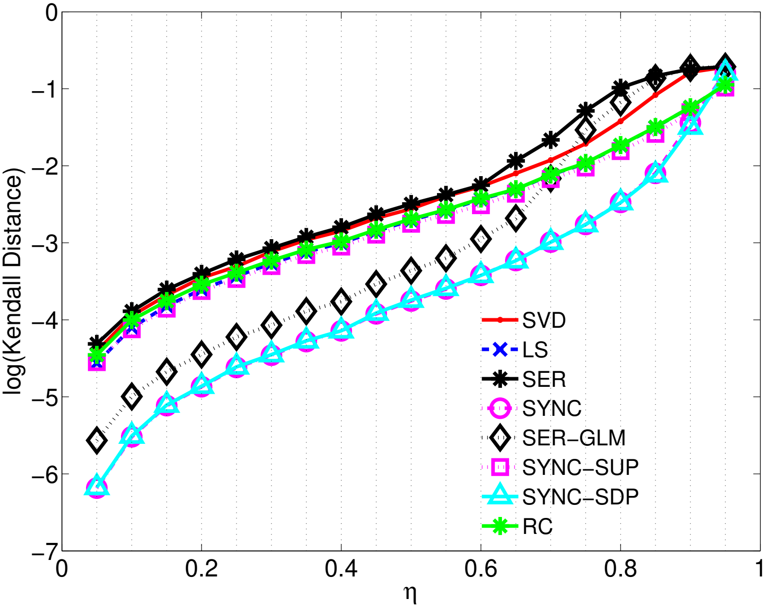
$$\frac{\vee \ [\pi_1(i) > \pi_1(j) \ \wedge \ \pi_2(i) < \pi_2(j)]\}|}{} = \frac{nr.flips}{\binom{n}{2}} \tag{24}$$

- we compute the Kendall distance on a logarithmic scale

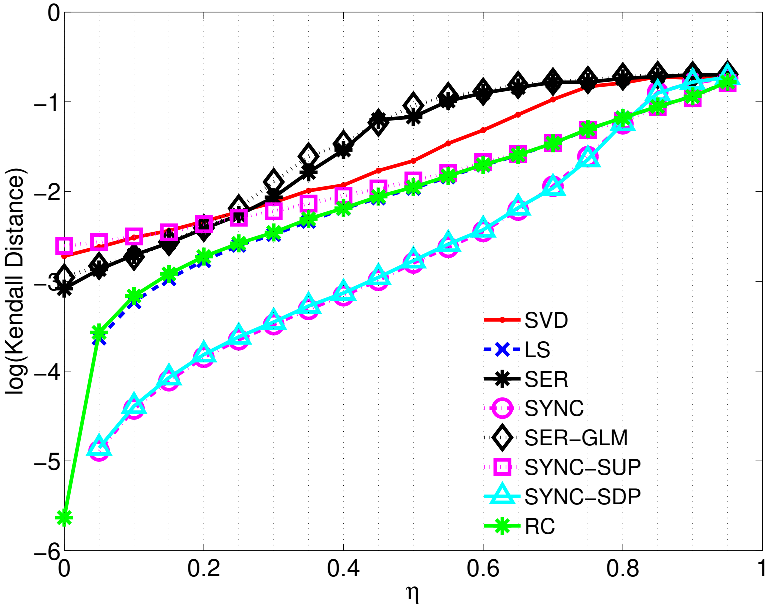
Let us look at recovery error, as we vary

- the noise level $\eta \in [0, 1)$ in the measurements
- the sparsity $p = \{1, 0.2, 0.05\}$ of the measurement graph.

Erdős-Rényi Outliers ERO($n = 200, p = 1, \eta$)

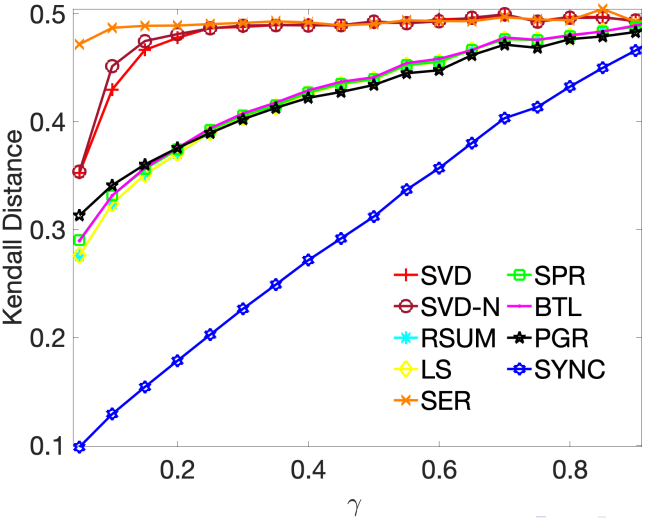


Erdős-Rényi Outliers ERO($n = 200, p = 0.2, \eta$)



Erdős-Rényi Outliers ERO($n = 1000, p = 0.05, \eta$)

- Sparse measurement graph
- Gamma-distributed players strengths/skills
- Comparison against additional state-of-the-art methods



FX matrix

Clean FX matrix:

	EUR	GBP	JPY	CHF	CAD	USD
EUR	-	0.84922 -08	120.621 +1.341	1.08374 +154	1.48042 +1312	1.08900 +220
GBP	1.17755 +11	-	142.036 +1.586	1.27613 +193	1.74322 +1562	1.28233 +273
JPY	0.008290 -93	0.007040 -80	-	0.008985 -87	0.012273 -28	0.009028 -83
CHF	0.92273 -131	0.78362 -119	111.300 +1.070	-	1.36601 +1008	1.00485 +65
CAD	0.67548 -604	0.57365 -519	81.478 +183	0.73206 -544	-	0.73560 -503
USD	0.91827 -186	0.77983 -166	110.763 +1.003	0.99518 -62	1.35944 +924	-

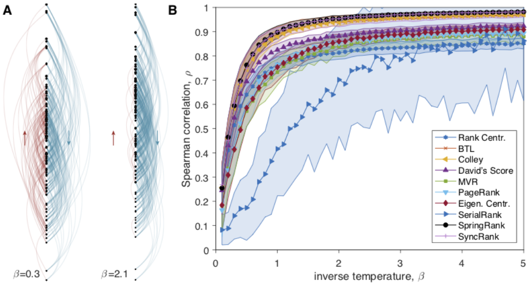
- Synchronize a perturbed FX matrix?
- Synchronize a perturbed beta matrix? Can synchronization be used to denoise an $n \times n$ matrix of pairwise betas? Can we better estimate the beta to the market?

A physical model for efficient ranking in networks

Caterina De Bacco,[1,2,*] Daniel B. Larremore,[3,4,2,†] and Cristopher Moore[2,‡]

[1] *Data Science Institute, Columbia University, New York, NY 10027, USA*
[2] *Santa Fe Institute, Santa Fe, NM 87501, USA*
[3] *Department of Computer Science, University of Colorado, Boulder, CO 80309, USA*
[4] *BioFrontiers Institute, University of Colorado, Boulder, CO 80303, USA*

We present a physically-inspired model and an efficient algorithm to infer hierarchical rankings of nodes in directed networks. It assigns real-valued ranks to nodes rather than simply ordinal ranks, and it formalizes the assumption that interactions are more likely to occur between individuals with similar ranks. It provides a natural statistical significance test for the inferred hierarchy, and it can be used to perform inference tasks such as predicting the existence or direction of edges. The ranking is obtained by solving a linear system of equations, which is sparse if the network is; thus the resulting algorithm is extremely efficient and scalable. We illustrate these findings by analyzing real and synthetic data, including datasets from animal behavior, faculty hiring, social support networks, and sports tournaments. We show that our method often outperforms a variety of others, in both speed and accuracy, in recovering the underlying ranks and predicting edge directions.

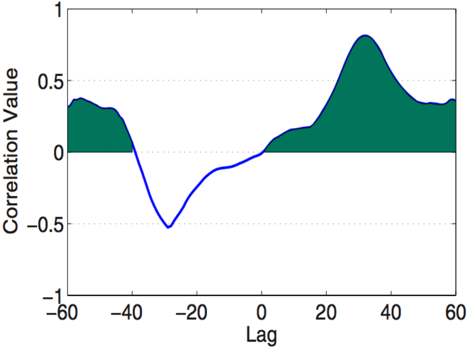


- C. De Bacco, D. B. Larremore and C. Moore, A physical model for efficient ranking in networks, Science Advances (2018). + Implementations in Python, Matlab, R.

Leaders and laggers in multivariate time series

- lead-lag networks from multivariate time series
- lagged relationships encountered in natural physical systems (correlation between two time series shifted in time relative to one another)
- one time series has a delayed response
  - to the other series,
  - to a common factor/stimulus that affects both series
- ranking/denoising arises from lagged relationships
- the return of instrument *i* on day *t* may influence the behavior of instrument *j* on day $t + 3$ (*i leads j* by 3 units of time)
- such pairwise comparisons are very noisy and inconsistent
- capture lead-lag relationships, compute rankings, predict the laggers catching up

Cross-correlations and the lead-lag matrix



Wu et al, 2010

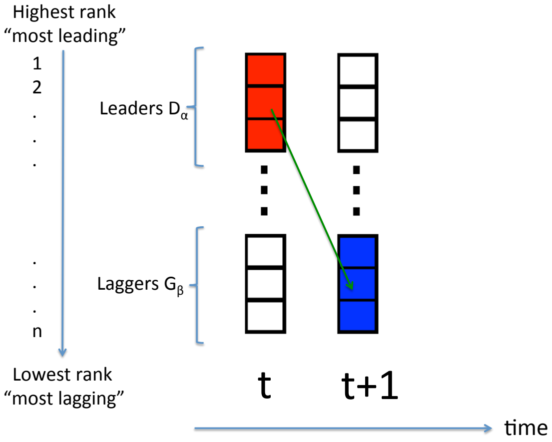
Options for building the pairwise comparison matrix:

1. C_{ij}: lag that maximizes the cross-correlation
2. C_{ij}: $\pm \max\{$avg. corr. of +ve lags, avg. corr. of -ve lags$\}$
3. C_{ij}: second order signatures of the two time series

$$A_{ij}(t-m,t) = \iint_{t-m<u<v<t} dX_i(u)\, dX_j(v) - dX_j(u)\, dX_i(v)$$

Global ranking of the time series

- find a global ranking of the time series,
- construct a leading and a lagging cluster,
- build a forecast for the lagging cluster catching up to the leading cluster



Data set: S&P 500 constituents (470)

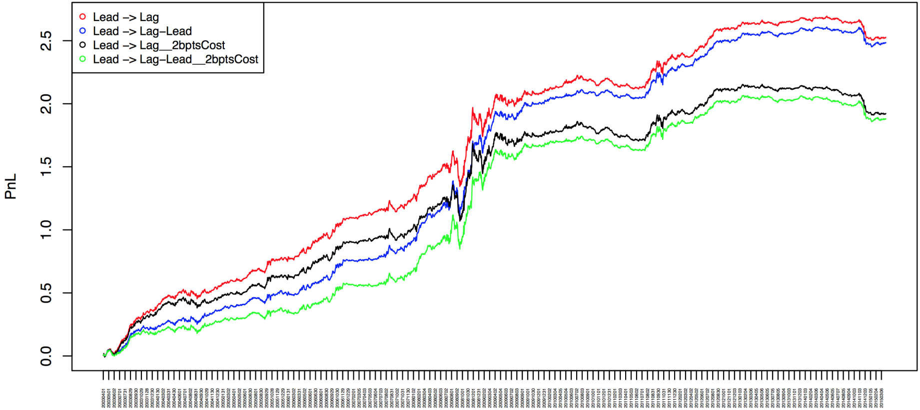
- 2003-2014, 3000+ trading days
- daily log returns
- on any given day, use the past $m = 60$ days of historical data
- forecast the future 1-day return
- universe given by S&P 500 constituents (and 10 sector ETFs)
- 1-to-1 hedge with $\{$SPY, the basket of leaders$\}$

Note: simple momentum-based approach ($D = G = [n]$) yields

- Sharpe = 0.5
- P&L = 1 bpts/day

Keep $\beta = 1 - \alpha$ fixed (could replace by a ”Cheeger sweep”)

Profit & Loss across time (2003-2014)



Ranking literature

- R. A. Bradley and M. E. Terry. *Rank analysis of incomplete block designs: I. the method of paired comparisons*. Biometrika, page 324345, 1952
- L. Page, S. Brin, R. Motwani, and T. Winograd. *The PageRank citation ranking: Bringing order to the Web*. In Proc. of the 7th Intl. World Wide Web Conference, pages 161–172, 1998
- David F Gleich and Lek-heng Lim. *Rank aggregation via nuclear norm minimization*. 17th ACM SIGKDD Intl conference on Knowledge discovery and data mining, pages60–68. ACM, 2011
- Sahand Negahban, Sewoong Oh, and Devavrat Shah. *Iterative ranking from pairwise comparisons*. In Advances in Neural Information Processing Systems 25, pages 2474–2482, 2012
- F. Fogel, A. d'Aspremont, and M. Vojnovic. *Serial-Rank: Spectral ranking using seriation*. In Advances in Neural Information Processing Systems 27, pages 900–908, 2014
- Fajwel Fogel, Alexandre d'Aspremont, and Milan Vojnovic. *Spectral ranking using seriation*. Journal of Machine Learning Research, 17(88):1–45, 2016.
- M. Cucuringu. *Sync-Rank: Robust Ranking, Constrained Ranking and Rank Aggregation via Eigenvector and Semidefinite Programming Synchronization*. IEEE Transactions on Network Science and Engineering, 3(1):58–79, 2016.
- Caterina De Bacco, Daniel B. Larremore, and Cristopher Moore. *A physical model for efficient ranking in networks*. Science Advances, 4(7), 2018
- A. d'Aspremont, M. Cucuringu, H. Tyagi, *Ranking and synchronization from pairwise measurements via SVD*, Journal of Machine Learning Research (JMLR), 22(19):1-63 (2021)