Dynamic ranking and translation synchronization

Eglantine Karlé

(joint work with Hemant Tyagi and Ernesto Araya)

Inria Lille, France

March 9, 2023

Introduction

- In many applications, we are given access to data involving pairwise comparisons between n items.
- Natural to model this as a graph (or graphs) with n vertices; edges denote the pairs of items that were compared.
- **Goal**: Estimate a latent signal from the pairwise comparison data.

Ranking

- Applications: sports tournaments, recommender systems etc.
- Pairwise information of the form "**item** *i* **beats item** *j*".
- **Goal**: Recover the latent ranking of the items.
    - Requires graph connectivity

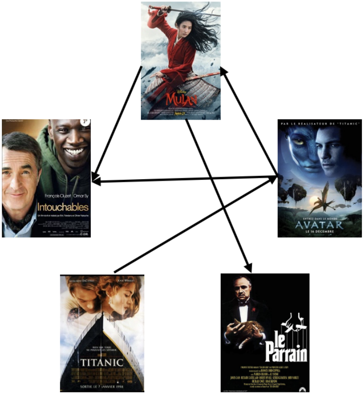


Figure: Movie *i* preferred over *j*

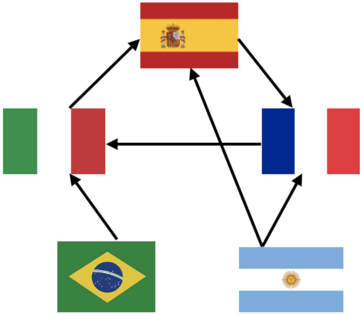
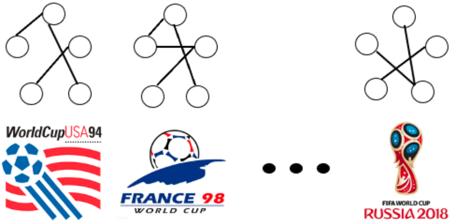


Figure: Team *i* beats *j*

Ranking in a dynamical setting

- Data is time-evolving in many applications. Example: Sports tournaments.
- We are given a sequence of graphs $G_0, G_1, \ldots, G_T$ and corresponding pairwise data for each graph.



- **Goal**: Estimate ranking at every time t.
- Note that each G_t may be very sparse, possibly disconnected ...
  - $\Longrightarrow$ Need assumptions on evolution model

Translation synchronization

- Let $z^* \in \mathbb{R}^n$ denote latent strengths of items.
- **Given:** Comparison graph $G = ([n], \mathcal{E})$ with data

$$y_{ij} = (z_i^* - z_j^*) + \varepsilon_{ij}, \quad \{i,j\} \in \mathcal{E}$$

  where ε_{ij} is i.i.d (0-mean) subgaussian noise.
- **Goal:** Estimate z^* (assuming G connected).
- Main applications: time synchronization of distributed networks, ranking.
- Well studied with ℓ_2 and ℓ_∞ error bounds (e.g., [HLBH17, dCT21])
- Can be viewed as an "approximation" of BTL model with $z_i^* = \log w_i^*$.

Dynamic Translation synchronization model [AKT22]

- $z_t^* \in \mathbb{R}^n$ is the strength vector at time $t \in \mathcal{T}$ (unif. grid on $[0, 1]$).
  - Recall $|\mathcal{T}| = T + 1$.
- **Observations**
  - Comparison graphs $G_t = ([n], \mathcal{E}_t)$ for all $t \in \mathcal{T}$.
  - Noisy measurements:

$$y_{ij}(t) = z^*_{t,i} - z^*_{t,j} + \varepsilon_{ij}(t) \quad \forall t \in \mathcal{T}, \forall \{i, j\} \in \mathcal{E}_t,$$

where ε is i.i.d subgaussian noise.

Goal

Recover z_t^* for all $t \in \mathcal{T}$.

- Each z_t^* assumed to be centered w.l.o.g., i.e., ${z_t^*}^\top \mathbf{1}_n = 0$.
- Individual graphs not necessarily connected.

Smoothness assumption

Assumption (Global Smoothness)

Let $C \in \mathbb{R}^{n \times \binom{n}{2}}$ to be the edge incidence matrix of the complete graph K_n. We assume that

$$\sum_{k=0}^{T-1} \left\| C^\top (z_k^* - z_{k+1}^*) \right\|_2^2 \leq S_T. \tag{0.1}$$

- Strength differences do not change too quickly on average.
  - $\iff$ z_k^* does not change quickly on average.
  - Interesting regime: $S_T = o(T)$
- Let $z^* = \begin{pmatrix} z_0^* \\ \vdots \\ z_T^* \end{pmatrix} \in \mathbb{R}^{n(T+1)}$. Can rewrite assumption as

$$\|Ez^*\|_2^2 \leq S_T.$$

  for a smoothness operator E.

1^{st} estimator : Penalized Least-Squares

- **Smoothness-penalized least-squares**

$$\hat{z} = \underset{\substack{z_0,\dots,z_T \in \mathbb{R}^n \\ z_k^\top \mathbf{1}_n = 0}}{\operatorname{argmin}} \sum_{t\in\mathcal{T}} \sum_{(i,j)\in\vec{\mathcal{E}}_t} (y_{ij}(t) - (z_{t,i} - z_{t,j}))^2 + \lambda \|Ez\|_2^2,$$

where $\hat{z} = \begin{pmatrix} \hat{z}_0 \\ \vdots \\ \hat{z}_T \end{pmatrix}$.

- Choose λ to balance bias-variance tradeoff.
  - Spectrum of operator E plays a key role.

2^{nd} estimator : Projection method

1. For each $t \in \mathcal{T}$, $\check{z}_t \in \mathbb{R}^n$ is the (min ℓ_2 norm) least-squares solution of
$$y_{ij}(t) = z_{t,i} - z_{t,j} \quad \forall \{i,j\} \in \mathcal{E}_t.$$
2. Form the vector $\check{z} = \begin{pmatrix} \check{z}_0 \\ \vdots \\ \check{z}_T \end{pmatrix}$.
3. $\hat{z}$: Projection of $\check{z}$ onto the span of $\approx T\sqrt{\tau}$ smallest eigenvectors of $E^\top E$.
4. Choose τ to balance bias-variance tradeoff.
   - Spectrum of operator E plays a key role.

Results

Theorem (Smoothness-penalized least squares)

Suppose that G_t is connected for each $t \in \mathcal{T}$. For $\lambda = (\frac{T}{S_T})^{2/5}$, it holds with probability larger than $1 - \delta$

$$\|\hat{z} - z^*\|_2^2 \leq T^{4/5} S_T^{1/5} \Psi_{\mathrm{LS}}(n, \sigma, \delta) + \Psi'_{\mathrm{LS}}(n, \sigma, \delta).$$

Here, $\Psi_{\mathrm{LS}}(\cdot)$ and $\Psi'_{\mathrm{LS}}(\cdot)$ are functions of the parameters of the problem (independent of T).

- Note that if $S_T = o(T)$,

$$\frac{\|\hat{z} - z^*\|_2^2}{|\mathcal{T}|} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \|\hat{z}_t - z_t^*\|_2^2 \xrightarrow[T \to \infty]{} 0.$$

- Connectivity of each G_t required in analysis for technical reasons.

Results

Theorem (Projection method)

Suppose that G_t is connected for each $t \in \mathcal{T}$. For $\tau = (\frac{S_T}{T})^{2/3}$, it holds with probability larger than $1 - \delta$ that

$$\|\hat{z} - z^*\|_2^2 \leq T^{2/3} S_T^{1/3} \Psi_{\text{Proj}}(n, \sigma, \delta) + \Psi'_{\text{Proj}}(n, \sigma, \delta)$$

where $\Psi_{\text{Proj}}(\cdot)$ and $\Psi'_{\text{Proj}}(\cdot)$ are functions of the parameters of the problem (independent of T).

- For $S_T = O(\frac{1}{T})$, we have $\|\hat{z} - z^*\|_2^2 = O(T^{1/3})$.
  - Matches rate for estimation of Lipschitz functions over a regular grid.
- Similar rate can be obtained for Penalized LS with an additional (non-intuitive) technical assumption...

Numerical results

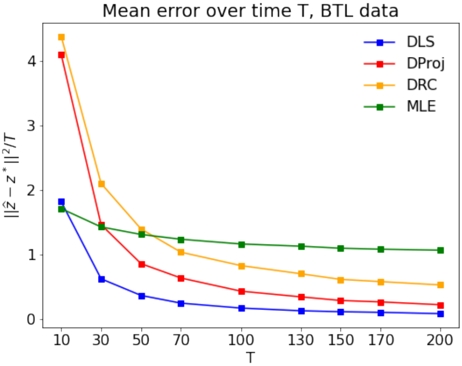


Figure: MSE for $n = 100$. Data are generated according to the Dynamic BTL model and the graphs are $G(n, p(t))$ with $p(t)$ randomly chosen in $[\frac{1}{n}, \frac{\log n}{n}]$. Graphs can be disconnected but the union graph is connected. We compare our estimators (DRC, DLS, DPro) with 'MLE' method [BLSR20].

Real data : English Premier League dataset

- **Data :** results of every game between seasons 2000/2001 and 2017/2018 for the $n = 43$ teams in EPL in this timeframe.
- **Individual graph :** average results for two successive seasons, resulting in $|\mathcal{T}| = 9$ graphs.
- **Performance criteria :**

$$\text{MSE} = \frac{1}{T+1} \sum_{t \in \mathcal{T}} \sum_{\{i,j\} \in \mathcal{E}_t} \left(y_{ij}(t) - (\hat{z}_{t,i} - \hat{z}_{t,j})\right)^2 .$$

$$\text{Number of upsets} := \sum_{t \in \mathcal{T}} \sum_{\{i,j\} \in \mathcal{E}_t} \mathbf{1}_{sign(y_{ij}(t)) \neq sign(\hat{z}_{t,i} - \hat{z}_{t,j})} .$$

	MSE	Upsets
Static LS	0.0024	0.67
DLS	0.0015	0.57
DProj	0.0014	0.58

Table: Performance for the Premier League dataset.

Summary and future directions

Summary

- Dynamic translation synchronization
  - Global smoothness assumption on the grid.
  - MSE bounds for latent strength recovery (over the grid).
  - Error vanishes as grid size increases.

Future research

- Consider models which incorporate dependence between observations over time.
  - Bigger picture: Learning dynamical processes over dynamical networks.
- Robustness to adversarial corruptions?

References I

- E. Araya, E. Karlé, and H. Tyagi, *Dynamic ranking and translation synchronization*, arXiv preprint arXiv:2207.01455 (2022).
- H. Bong, W. Li, S. Shrotriya, and A. Rinaldo, *Nonparametric estimation in the dynamic Bradley-Terry model*, AISTATS, 2020, pp. 3317–3326.
- Alexandre d'Aspremont, Mihai Cucuringu, and Hemant Tyagi, *Ranking and synchronization from pairwise measurements via SVD*, J. Mach. Learn. Res. **22** (2021), 19:1–19:63.
- Xiangru Huang, Zhenxiao Liang, Chandrajit Bajaj, and Qixing Huang, *Translation synchronization via truncated least squares*, Advances in neural information processing systems **30** (2017).