Sharp threshold for alignment of graph databases with Gaussian weights

Luca Ganassali

Mathematical and Scientific Machine Learning, August 16-19th, 2021

INRIA, DI/ENS, PSL Research University, Paris, France

Graph/matrix alignment

Question: Given two graphs $G = (V, E)$ and $G' = (V', E')$ with $|V| = |V'|$, *what is the best way to match nodes of G with nodes of G'?*

Question: Given two graphs $G = (V, E)$ and $G' = (V', E')$ with $|V| = |V'|$, *what is the best way to match nodes of G with nodes of G'?*

Minimizing disagreements: Find a bijection $f : V \to V'$ that minimizes

$$\sum_{(i,j)\in V^2} \left(\mathbf{1}_{(i,j)\in E} - \mathbf{1}_{(f(i),f(j))\in E'}\right)^2,$$

or, equivalently solve

$$\max_{\Pi} \langle G, \Pi G' \Pi^\top \rangle,$$

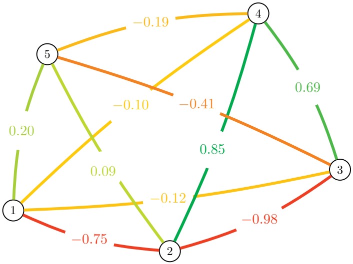
where Π runs over all permutation matrices.

Correlated Wigner model:

- Draw the planted permutation π^* uniformly at random in $\mathcal{S}_n$.
- Set

$$B = \rho \cdot {\Pi^*}^T A \Pi^* + \sqrt{1-\rho^2} \cdot H,$$

where H is an independent copy of A, and $\Pi^*_{i,j} = \mathbf{1}_{j=\pi^*(i)}$.



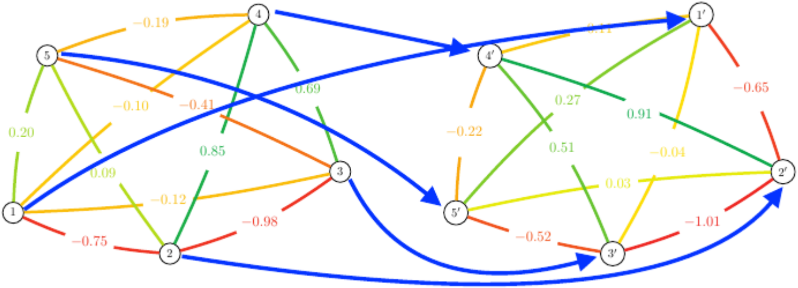
Planted Alignment with gaussian weights

Correlated Wigner model:

- Draw the planted permutation π^* uniformly at random in $\mathcal{S}_n$.
- Set

$$B = \rho \cdot {\Pi^*}^T A \Pi^* + \sqrt{1-\rho^2} \cdot H,$$

where H is an independent copy of A, and $\Pi^*_{i,j} = \mathbf{1}_{j=\pi^*(i)}$.



Main result: sharp threshold for exact alignment at $n\rho^2 / \log n \sim 4$

Theorem (Achievability part)

If for n large enough

$$\rho^2 \geq \frac{(4+\varepsilon)\log n}{n} \tag{1}$$

for some $\varepsilon > 0$, then there is an estimator (namely, the MAP estimator) $\hat{\pi}$ of π given A, B such that $\hat{\pi} = \pi^*$ with probability $1 - o(1)$.

Theorem (Converse part)

Conversely, if

$$\rho^2 \leq \frac{4\log n - \log\log n - \omega(1)}{n} \tag{2}$$

then any estimator $\hat{\pi}$ of π given A, B verifies $\hat{\pi} = \pi^*$ with probability $o(1)$.