Matching Embeddings via Shuffled Total Least Squares Regression

Daniel L. Sussman and Qian Wang

Boston University

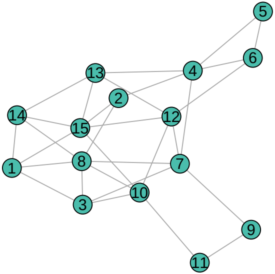
June 29, 2023

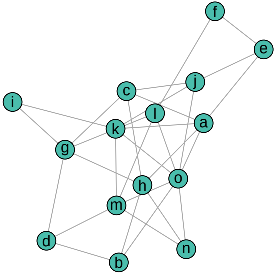
What is graph matching?

Formulation

Consider observing two graphs, $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$.
The classical graph matching formulation is to find a map $\pi : V_1 \mapsto V_2$, that minimizes the symmetric difference between

$$\pi(E_1) = \{(\pi(i), \pi(j)) : (i, j) \in E_1\} \text{ and } E_2.$$



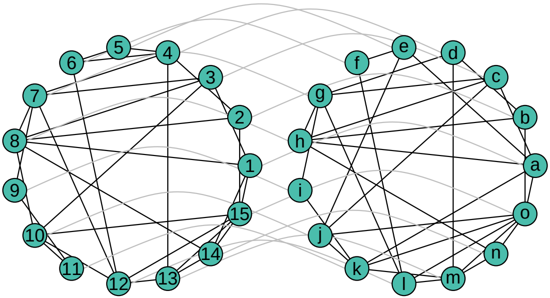


What is graph matching?

Formulation

Consider observing two graphs, $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$.
The classical graph matching formulation is to find a map $\pi : V_1 \mapsto V_2$, that minimizes the symmetric difference between

$$\pi(E_1) = \{(\pi(i), \pi(j)) : (i, j) \in E_1\} \text{ and } E_2.$$

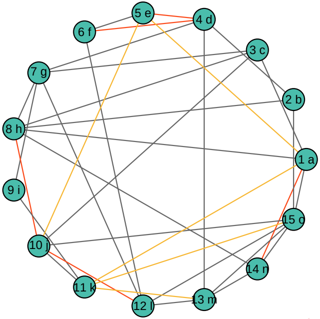


What is graph matching?

Formulation

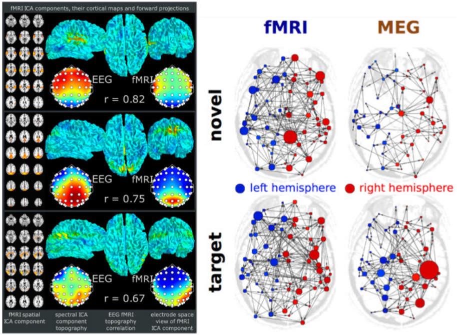
Consider observing two graphs, $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$.
The classical graph matching formulation is to find a map $\pi : V_1 \mapsto V_2$, that minimizes the symmetric difference between

$$\pi(E_1) = \{(\pi(i), \pi(j)) : (i, j) \in E_1\} \text{ and } E_2.$$



Should we do graph matching?

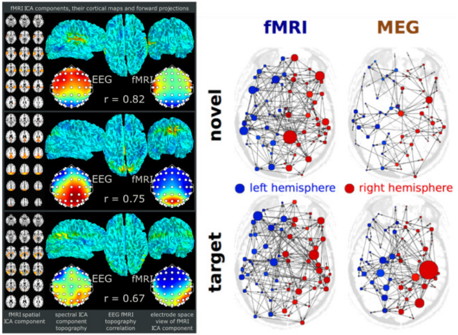
- Matching graphs from different modalities



Calhoun and Sui 2016

Should we do graph matching?

- Matching graphs from different modalities

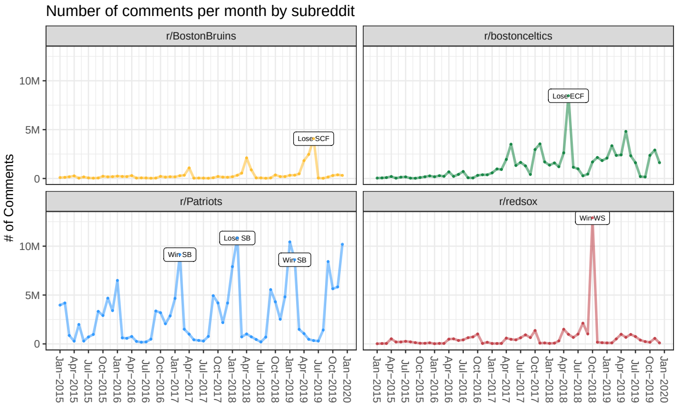


Calhoun and Sui 2016

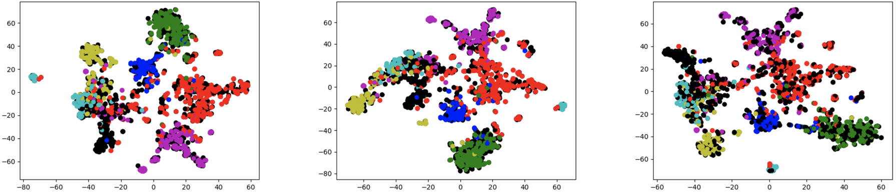
- Matching a social network to a co-purchasing network

Should we do graph matching

- Matching across topics and time periods



A first step: match embeddings



Zhang et al. 2019

Idea

1. Represent graphs as point clouds
2. Align point clouds

Shuffled Linear Regression - Model

- Model:

$$Y = \Pi^* XR + E,$$

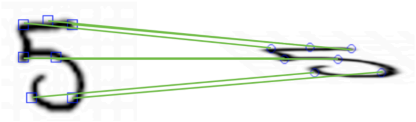
where $\Pi^* \in \mathcal{P}_n$ is an unknown permutation matrix.

  - Pananjady et al. 2016; Pananjady et al. 2017; Flammarion et al. 2016; Collier and Dalalyan 2016

Shuffled Linear Regression - Applications

Pose and correspondence estimation

- Goal: find similar objects across images from different perspectives



(Pananjady et al. 2017)

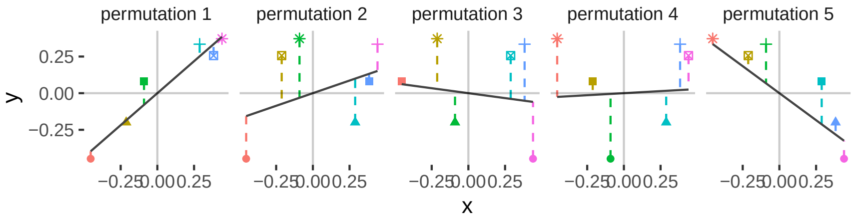
Header-free communication

- Goal: Recover signal origins without without sending location information

Shuffled Linear Regression - Estimation

- OLS estimate:

$$(\hat{\Pi}, \hat{R}) = \arg \min_{\Pi \in \mathcal{P}_n, R \in \mathbb{R}^{p \times p}} \|Y - \Pi X R\|_F^2.$$



A more realistic model

Model:

$$Y_1 = X + E_1$$
$$Y_2 = \Pi^* X R + E_2, \tag{1}$$

- Permutation: $\Pi^* \in \mathcal{P}_n$
- Design: $X \in \mathbb{R}^{n \times p}$
- Coefficient: $R \in \mathbb{R}^{p \times p}$
- Noise: $E_1, E_2 \in \mathbb{R}^{n \times p}$

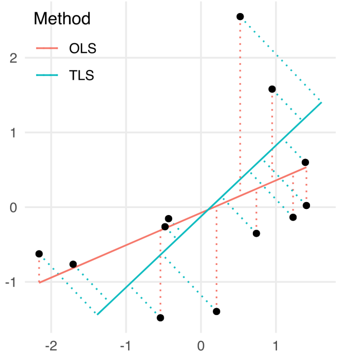
- Given the observations (Y_1, Y_2), estimate Π^*.

Total Least Squares (TLS) Estimator

The TLS estimator for *errors-in-variables regression*:

$$\min_{\hat{Y}_1, \hat{Y}_2 \in \mathbb{R}^{p\times p}} \|[Y_2 | Y_1] - [\hat{Y}_2 | \hat{Y}_1]\|_F^2$$

$$s.t. \ \mathrm{rank}([\hat{Y}_2 | \hat{Y}_1]) \le p.$$



Shuffled TLS Estimator

Let

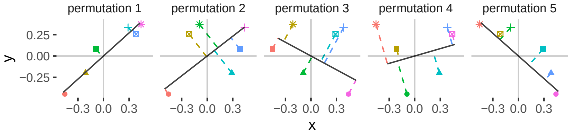
$$Y_\Pi = [Y_2 | \Pi Y_1], \quad M_\Pi = [\Pi^* X R | \Pi X], \text{ and } \quad E_\Pi = [E_2 | \Pi E_1]$$

Write model (1) as

$$Y_\Pi = M_\Pi + E_\Pi, \tag{2}$$

The shuffled TLS estimator is

$$\hat{\Pi} = \arg\min_{\Pi \in \mathcal{P}_n} \sum_{i=p+1}^{2p} \sigma_i^2(Y_\Pi). \tag{3}$$



Evaluation Method and Identifiability Issue

- The Hamming distance

$$d_H(\hat{\Pi}, \Pi^*) = \#\{i | \hat{\Pi}(i) \neq \Pi^*(i)\},$$

- The normalized quadratic loss

$$\frac{1}{np} \|\hat{\Pi} X - \Pi^* X\|_F^2.$$

Evaluation Method and Identifiability Issue

Example (Identifiability Issue of the Shuffled TLS Estimator)

Consider a noiseless case when $E_1 = E_2 = 0$, $\Pi^* = I_n$, $R = I_p$, *let* $n = 10$, *and*

$$Y_1 = Y_2 = X = \begin{bmatrix} \mathbf{1}_5 & -\mathbf{1}_5 \\ \mathbf{1}_5 & \mathbf{1}_5 \end{bmatrix}.$$

Evaluation Method and Identifiability Issue

Example (Identifiability Issue of the Shuffled TLS Estimator)

Consider a noiseless case when $E_1 = E_2 = 0$, $\Pi^* = I_n$, $R = I_p$, *let* $n = 10$, *and*

$$Y_1 = Y_2 = X = \begin{bmatrix} \mathbf{1}_5 & -\mathbf{1}_5 \\ \mathbf{1}_5 & \mathbf{1}_5 \end{bmatrix}.$$

- Normalized Proscrustes quadratic loss:

$$\frac{1}{\|X\|_F^2} \min_{Q \in \mathcal{O}(p)} \|\Pi^* X - \hat{\Pi} X Q\|_F^2.$$

Lemma

Assume the condition number $\kappa(X) = 1$, *we have the relationship*

$$\min_{Q \in \mathcal{O}(p)} \|\Pi^* X - \Pi X Q\|_F^2 \leq 2 \sum_{i=1+p}^{2p} \sigma_i^2(\Pi^* X | \Pi X).$$

Model assumptions and Main Result

We assume the following conditions hold:

Assumption (Design Matrix)

The latent design matrix has condition number $\kappa(X) = 1$.

Assumption (Coefficient Matrix)

$\sigma_p(R) \leq 1$ *and* $\sigma_1(R) \geq 1$.

Assumption (Noise Variables)

$E_{1i}, E_{2i} \sim^{i.i.d} N(0, \Sigma)$ *for* $i \in [n]$.

Main Result

Theorem

For the statistical model (1), *under the assumptions, the total least squares estimator* $\hat{\Pi}$ *satisfies*

$$\frac{\min_{Q\in\mathcal{O}(p)} \|\Pi^* X - \hat{\Pi} X Q\|_F^2}{\|X\|_F^2} \leq \frac{4\lambda_1(\Sigma)}{\sigma_p^2(R)}(1+\eta a_n)\left[8\sqrt{2}\sigma_1(R)\frac{p\sqrt{n}}{\|X\|_F} + \frac{np}{\|X\|_F^2}\right], \quad (4)$$

where $a_n = \sqrt{\frac{\mathrm{tr}(\Sigma)}{\lambda_1(\Sigma)}\frac{\log(n)}{cn}}$, *with probability greater than*

$$1 - n^{-\eta^2},$$

where c *is at least* $\frac{1}{32}$.

Main Result

Define the signal-to-noise ratio as snr $= \frac{\|X\|_F^2/n}{\mathrm{tr}(\Sigma)}$.
The upper bound is approximately

$$c_1(R)\sqrt{\frac{\mathrm{tr}(\Sigma)}{\mathrm{snr}}} + c_2(R)\frac{1}{\mathrm{snr}}$$

- $X_{ij} \sim N(0,1), E_{ij} \sim N(0,\sigma^2)$, snr$\sim \frac{1}{\sigma^2}$

  $$c\sigma^2,$$

  where $c = c_1(R)\sqrt{p} + c_2(R)$.
- For the Procrustes loss to go to zero, snr needs to go to infinity.

Result Comparison

- Our bound:

$$c\sigma^2$$

- Pananjady et al. 2017

$$Y = \Pi^* X R^* + E,$$

- For $p < \log(n)$:

$$\frac{1}{np}\|\hat{\Pi} X \hat{R} - \Pi^* X R^*\|_F^2 \leq c_1 \sigma^2 (\frac{p}{n} + 1).$$

Result Comparison

- Our bound:

$$c\sigma^2$$

- Flammarion et al. 2016

$$Y = \Pi^* X^* + E,$$

where the columns of X^* is unimodal.

- $$\frac{1}{np}\|\hat{\Pi}\hat{X} - \Pi^* X^*\|_F^2 \leq \sigma^2(1 + \frac{\log(n)}{p}).$$

Permutation Recovery in Shuffled Linear Regression is NP-hard

$$
\begin{aligned}
&\min_{\Pi} \min_{R} \|\Pi X R - Y\|_F^2 \\
&= \min_{\Pi} \|\Pi X (X^T X)^{-1} X^T \Pi^T Y - Y\|_F^2 \\
&= \min_{\Pi} \operatorname{tr}(\Pi (Z^T Z - 2Z) \Pi^T Y Y^T),
\end{aligned}
$$

where $Z = X(X^T X)^{-1} X^T$.

- Shuffled OLS is equivalent to a QAP

The Alternating LAP/OLS Algorithm (ALOA)

Model:

$$Y_2 = \Pi^* Y_1 R + E_2$$

Algorithm:
Iterate between

- Step 1: given $\hat{\Pi}$, estimate R via OLS.
- Step 2: given $\hat{R}$, estimate $\hat{\Pi}$ by solving a LAP, assigning the n rows of Y_2 to the n rows of $Y_1 \hat{R}$.

$$C_{ij} = \| Y_{2i} - (Y_1 \hat{R})_j \|_F^2.$$

The Alternating LAP/TLS Algorithm, ALTA

Model:

$$Y_1 = X + E_1, Y_2 = \Pi^* X R + E_2$$

Algorithm:

Iterate between

- Step 1: given $\hat{\Pi}$, estimate (X, R) via TLS.
- Step 2: given $(\hat{X}, \hat{R})$ ... ?

The Alternating LAP/TLS Algorithm, ALTA

Model:

$$Y_1 = X + E_1, Y_2 = \Pi^* XR + E_2$$

Algorithm:

Iterate between

- Step 1: given $\hat{\Pi}$, estimate (X, R) via TLS.
- Step 2: given $(\hat{X}, \hat{R})$ ... ?
    - $\arg\min_{\Pi \in \mathcal{P}_n} \sum_{i=p+1}^{2p} \sigma_i^2([Y_2 | \Pi Y_1])$, does not depend on $(\hat{X}, \hat{R})$.

The Alternating LAP/TLS Algorithm, ALTA

Model:

$$Y_1 = X + E_1,\ Y_2 = \Pi^* XR + E_2$$

Algorithm:

Iterate between

- Step 1: given $\hat{\Pi}$, estimate (X, R) via TLS.
- Step 2: given $(\hat{X}, \hat{R})$ ... ?
  - $\arg\min_{\Pi \in \mathcal{P}_n} \sum_{i=p+1}^{2p} \sigma_i^2([Y_2 | \Pi Y_1])$, does not depend on $(\hat{X}, \hat{R})$.

How do we define a LAP?

- ALTA_1:

$$C_{ij}^{(1)} = \| Y_{2i} - \hat{R}^T \hat{X}_j \|_F^2 + \| Y_{1j} - \hat{X}_i \|_F^2$$

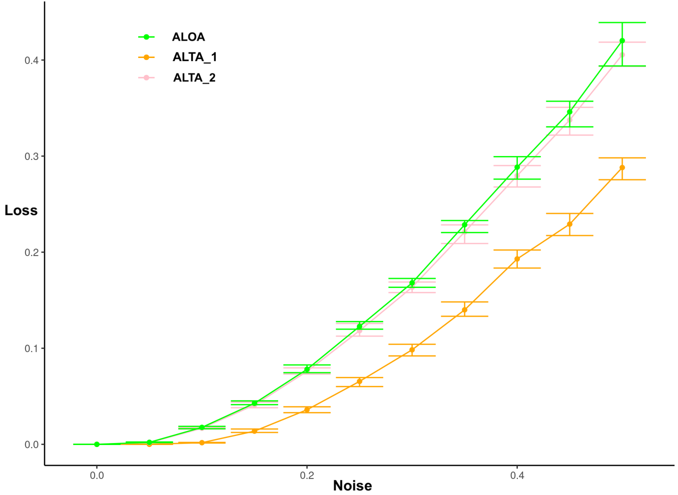
.

- ALTA_2:

$$C_{ij}^{(2)} = \min_{x \in \mathbb{R}^d} \| Y_{2i} - \hat{R}^T x \|_F^2 + \| Y_{1j} - x \|_F^2$$

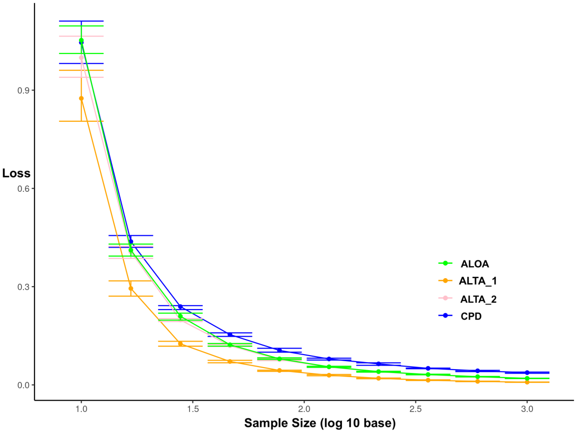
Simulation Studies

Initiate all algorithms at $\Pi = I_n$.



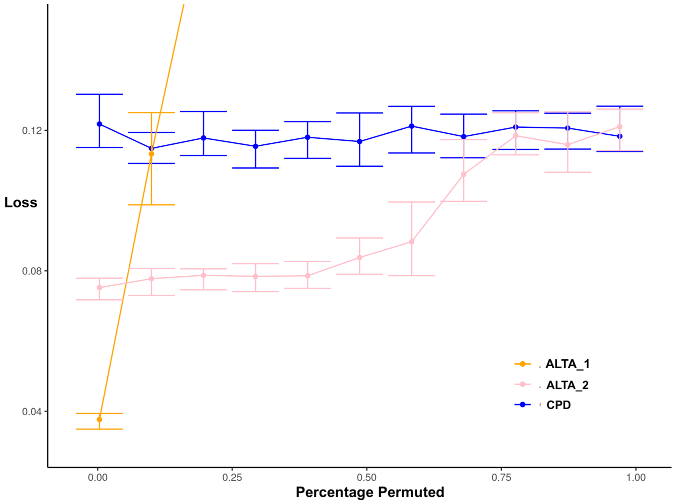
Simulation Studies

Increase the signal-to-noise ratio via decreasing the noise like $\frac{1}{n}$. (CPD, (Myronenko and Song 2010).)



Simulation Studies

Initialize further away from the truth.



Contributions

- Propose estimate $\hat{\Pi}$ based on the TLS method.
- Provide an upper bound on the Procrustes quadratic loss.
    - Many works in the shuffled linear regression setting, less so in the shuffled TLS regression.
    - Perhaps due to the difficulty in analyzing singular values compared with Frobenius norm.
- Approximate $\hat{\Pi}$ via ALTA.
    - The permutation recovery problem continuous to be an open challenge to researchers of various fields.

Potential Extension and Future Research

- Relax the assumptions:
  - $\kappa(X) = 1$.

$$Y_1 = XR_1 + E_1$$
$$Y_2 = \Pi^* XR_2 + E_2$$

    - Allow $\dim(R_1) \neq \dim(R_2)$?
  - $E_{1i}, E_{2i} \sim^{i.i.d} N(0, \Sigma)$.
    1. $E_{1i} \sim N(0, \Sigma_1), E_{2i} \sim N(0, \Sigma_2)$
    2. E_1 correlated with E_2 (This can happen when the two graphs A and B are correlated.)
- Extend theorem to big p, say, $p > \log(n)$.

References I

Calhoun, Vince D and Jing Sui (2016). "Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness". In: *Biological psychiatry: cognitive neuroscience and neuroimaging* 1.3, pp. 230–244.

Collier, O. and A. S. Dalalyan (2016). "Minimax rates in permutation estimation for feature matching". In: *J. Mach. Learn. Res., vol. 17, no. 1, pp. 162–192.*

Flammarion, N. et al. (2016). "Optimal rates of statistical seriation.". In: *Online. Available at https://arxiv.org/abs/1607.02435.*

Myronenko, A. and X. Song (2010). "Point-Set Registration: Coherent Point Drift". In: *IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, no. 12, pp. 2262-2275, Dec.*

References II

- Pananjady, Ashwin et al. (2016). "Linear regression with an unknown permutation: Statistical and computational limits". In: *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pages 417–424.*
- — (2017). "Denoising linear models with permuted data". In: *Information Theory (ISIT), 2017 IEEE International Symposium on, pages 446–450. IEEE.*
- Zhang, Si et al. (2019). "Origin: Non-rigid network alignment". In: *2019 IEEE International Conference on Big Data (Big Data).* IEEE, pp. 998–1007.

Thanks!

Questions?

Acknowledgement

This material is based on research sponsored by the Air Force Research Laboratory and Defense Advanced Research Projects Agency (DARPA) under agreement number FA8750-20-2-1001. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory and DARPA, or the U.S. Government.