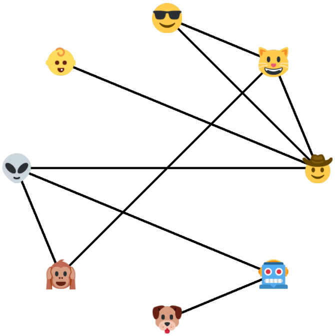
Graph matching from a statistical perspective

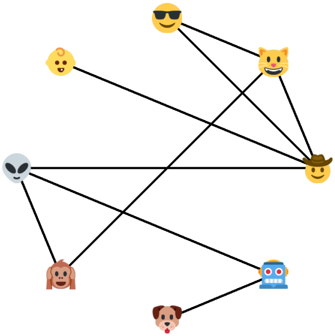
Jesús Arroyo

October 4th, 2023

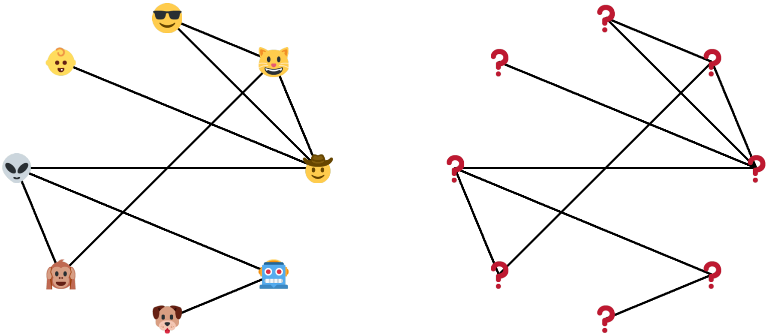
Texas A&M University
Stat Café

Graph matching

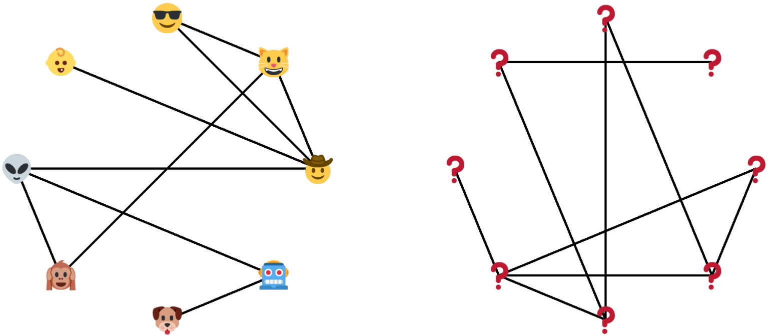




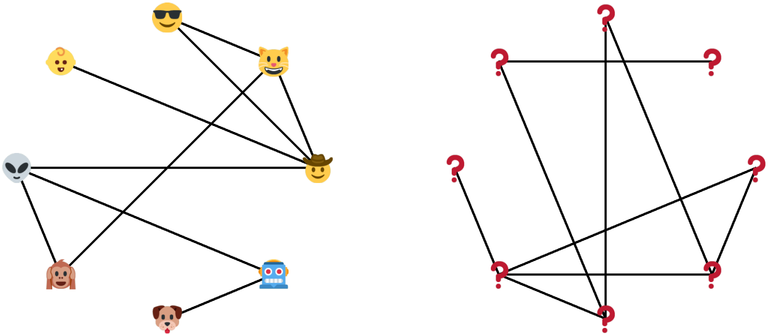
Graph matching



Graph matching



Graph matching

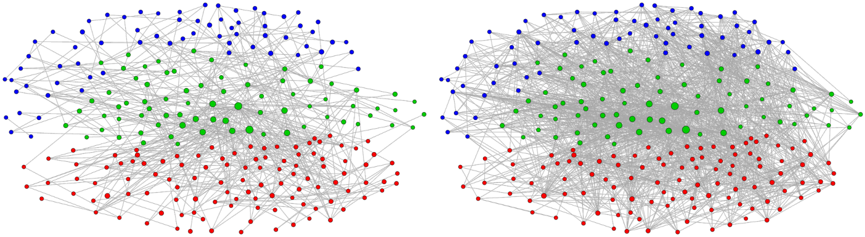


Exact graph matching = graph isomorphism problem

Graph matching

In real datasets, graph matching is usually **inexact**:

- Aligning biological networks
- Image/video/text processing
- De-anonymizing social networks
- Record linkage

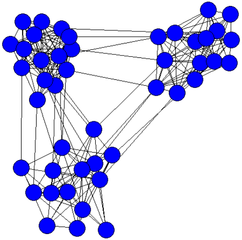


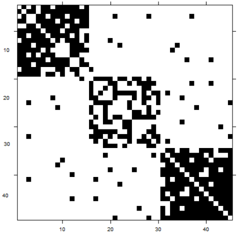
The C. elegans chemical and electrical connectomes (Chen et al. 2016, *Worm*)

Notation:

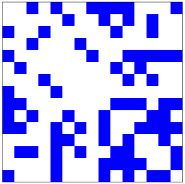
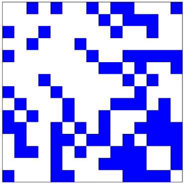
- Consider two *simple* graphs with n vertices each.
- Graphs are represented by their adjacency matrices $A, B \in \{0,1\}^{n \times n}$.

Goal: align the rows and columns of A and B





Graph matching problem

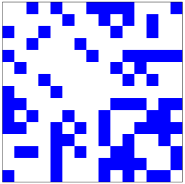
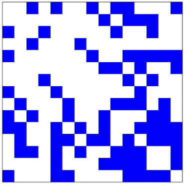
 

Two main approaches:

- Algorithmic: optimization, search strategies, spectral methods, etc. (Conte et al., 2004; Foggia et al., 2014). E.g.: *quadratic assignment problem* (QAP):

$$\underset{\text{permutation}\,P}{\operatorname{argmin}} \sum_{i \neq j} (A_{ij} - (PBP^\top)_{ij})^2 = \underset{\text{permutation}\,P}{\operatorname{argmax}} \left\langle A, PBP^\top \right\rangle .$$

Graph matching problem

Two main approaches:

- Algorithmic: optimization, search strategies, spectral methods, etc. (Conte et al., 2004; Foggia et al., 2014). E.g.: *quadratic assignment problem* (QAP):

$$\underset{\text{permutation}\,P}{\operatorname{argmin}} \sum_{i \neq j} (A_{ij} - (PBP^\top)_{ij})^2 = \underset{\text{permutation}\,P}{\operatorname{argmax}} \left\langle A, PBP^\top \right\rangle.$$

- Random graph models: pair of graphs generated from some distribution. E.g. correlated Erdős-Rényi graph model (Lyzinski et al., 2014)

$$A_{ij} \sim \mathsf{Ber}(p), \qquad B_{ij} \sim \mathsf{Ber}(p),$$

$$\mathsf{Corr}(A_{ij}, B_{ij}) = \rho \geq 0.$$

This talk

Overview:

- Random graph models for the graph matching problem
- Matching via maximum likelihood estimation
- Theory: when is MLE consistent for graph matching?
- Computational aspects: non-convex relaxations
- Illustrations on simulated and real networks

This talk

Overview:

- Random graph models for the graph matching problem
- Matching via maximum likelihood estimation
- Theory: when is MLE consistent for graph matching?
- Computational aspects: non-convex relaxations
- Illustrations on simulated and real networks

Problems considered:

- Unipartite to unipartite graph matching
- Bipartite to unipartite graph matching
- Some future directions

Outline

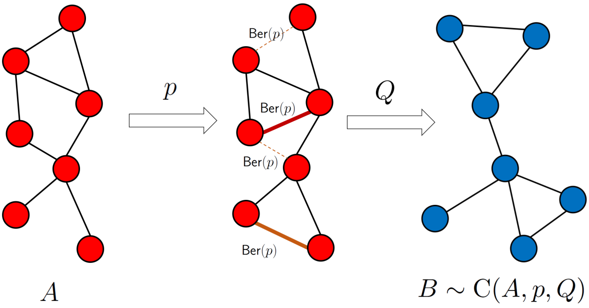
Graph matching in errorfully observed networks

Graph matching between bipartite and unipartite networks

Corrupting channel model

Model: B is an edge and vertex-label corrupted version of A

1. Flip edges and non-edges of A with probability p.
2. Shuffle vertices with permutation Q.



Maximum likelihood estimation and graph matching

- Maximum likelihood estimator:

$$(\widehat{p}_{\mathsf{MLE}}, \widehat{Q}_{\mathsf{MLE}}) := \underset{p,Q}{\operatorname{argmax}} \sum_{u>v} \log \mathbb{P}_p \left(A_{uv} = (QBQ^T)_{uv} \right).$$

Maximum likelihood estimation and graph matching

- Maximum likelihood estimator:

$$(\widehat{p}_{\mathsf{MLE}}, \widehat{Q}_{\mathsf{MLE}}) := \operatorname*{argmax}_{p,Q} \sum_{u>v} \log \mathbb{P}_p \left(A_{uv} = (QBQ^T)_{uv} \right).$$

- **Result**: MLE is equivalent to the QAP formulation:

$$\widehat{Q}_{\mathsf{MLE}} = \operatorname*{argmin}_{Q \in \Pi_n} \|A - QBQ^T\|_F^2.$$

Maximum likelihood estimation and graph matching

- Maximum likelihood estimator:

$$(\widehat{p}_{\mathsf{MLE}}, \widehat{Q}_{\mathsf{MLE}}) := \underset{p,Q}{\operatorname{argmax}} \sum_{u>v} \log \mathbb{P}_p \left(A_{uv} = (QBQ^T)_{uv} \right).$$

- **Result**: MLE is equivalent to the QAP formulation:

$$\widehat{Q}_{\mathsf{MLE}} = \underset{Q \in \Pi_n}{\operatorname{argmin}} \|A - QBQ^T\|_F^2.$$

- Extensions to non-uniform corrupting probabilities: the equivalence between MLE and QAP also holds.

When is the MLE correct?

- Difficulty lies on how different A and QAQ^T are, for any $Q \neq I$:

$$\|A - QAQ^T\|_F^2 = \sum_{i \neq j} (A_{ij} - A_{\sigma(i),\sigma(j)})^2.$$

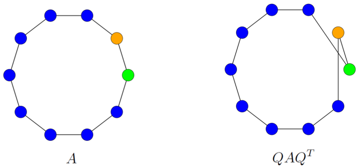
When is the MLE correct?

- Difficulty lies on how different A and QAQ^T are, for any $Q \neq I$:

$$\|A - QAQ^T\|_F^2 = \sum_{i \neq j} (A_{ij} - A_{\sigma(i),\sigma(j)})^2.$$

Examples:

- $k = 2$
- $\|A - QAQ^T\|_F^2 = 4.$



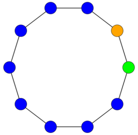
When is the MLE correct?

- Difficulty lies on how different A and QAQ^T are, for any $Q \neq I$:

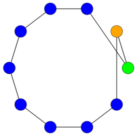
$$\|A - QAQ^T\|_F^2 = \sum_{i \neq j} (A_{ij} - A_{\sigma(i),\sigma(j)})^2.$$

Examples:

- $k = 2$
- $\|A - QAQ^T\|_F^2 = 4.$

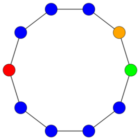


A

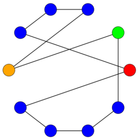


QAQ^T

- $k = 3$
- $\|A - QAQ^T\|_F^2 = 10.$



A



QAQ^T

$\Pi_{n,k}$ permutations that shuffle exactly k vertices.

Consistency of the MLE

Sequence of networks $\{A_n\}$ and parameters $\{p_n, Q_n\}$

Theorem (A., Sussman, Priebe, Lyzinski, 2021)

- $\widehat{Q}_{MLE}$ *is* **consistent** *(correct matching in the limit) if*

$$\min_{Q \in \Pi_{n,k}} \|A_n - QA_nQ^T\|_F^2 \geq \frac{6k \log n}{(1/2 - p_n)^2}, \qquad \forall k \geq 2.$$

Consistency of the MLE

Sequence of networks $\{A_n\}$ and parameters $\{p_n, Q_n\}$

Theorem (A., Sussman, Priebe, Lyzinski, 2021)

- $\widehat{Q}_{MLE}$ *is* **consistent** *(correct matching in the limit) if*

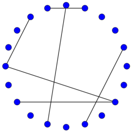
$$\min_{Q \in \Pi_{n,k}} \|A_n - QA_nQ^T\|_F^2 \geq \frac{6k \log n}{(1/2 - p_n)^2}, \qquad \forall k \geq 2.$$

- $\widehat{Q}_{MLE}$ *is* **not consistent** *if there exists* $m = \Omega(n)$ *disjoint permutations* $Q_1, \ldots, Q_m$

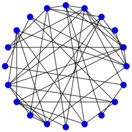
$$\max_{i \in [m]} \|A_n - Q_iA_nQ_i^T\|_F^2 = o\left(\frac{\log n}{(1/2 - p_n)^2}\right).$$

Erdös-Rényi graph model

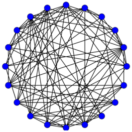
$A_n \sim \mathrm{G}(n, \alpha_n)$



$\alpha = 0.05$



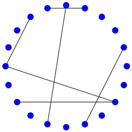
$\alpha = 0.2$



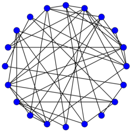
$\alpha = 0.4$

Erdös-Rényi graph model

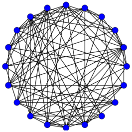
$A_n \sim \mathrm{G}(n, \alpha_n)$



$\alpha = 0.05$



$\alpha = 0.2$



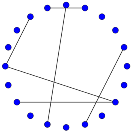
$\alpha = 0.4$

- $\widehat{Q}_{\mathrm{MLE}}$ is **consistent** if

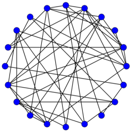
$$\alpha_n \geq c\sqrt{\frac{\log n}{n(1/2 - p_n)^2}}.$$

Erdös-Rényi graph model

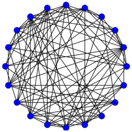
$A_n \sim \mathrm{G}(n, \alpha_n)$



$\alpha = 0.05$



$\alpha = 0.2$



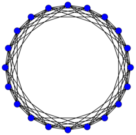
$\alpha = 0.4$

- $\widehat{Q}_{\mathrm{MLE}}$ is **consistent** if

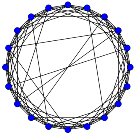
$$\alpha_n \geq c\sqrt{\frac{\log n}{n(1/2 - p_n)^2}}.$$

Small-world networks

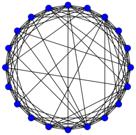
Newman-Watts model: $A_n \sim \mathrm{NW}(n, d_n, \beta_n)$,



$d = 4$,
$\beta = 0$



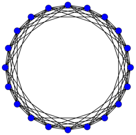
$d = 4$,
$\beta = 0.05$



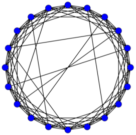
$d = 4$,
$\beta = 0.2$

Small-world networks

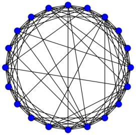
Newman-Watts model: $A_n \sim \mathrm{NW}(n, d_n, \beta_n)$,



$d = 4$,
$\beta = 0$



$d = 4$,
$\beta = 0.05$



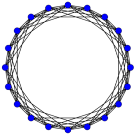
$d = 4$,
$\beta = 0.2$

- $\widehat{Q}_{\mathsf{MLE}}$ is **not consistent** if $(1/2 - p_n)^2 = o(\sqrt{\log n / n})$ and

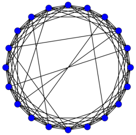
$$\beta_n = o\left(\frac{\log n}{(1/2 - p_n)^2 n}\right).$$

Small-world networks

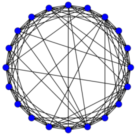
Newman-Watts model: $A_n \sim \mathrm{NW}(n, d_n, \beta_n)$,



$d = 4$,
$\beta = 0$



$d = 4$,
$\beta = 0.05$



$d = 4$,
$\beta = 0.2$

- $\widehat{Q}_{\mathsf{MLE}}$ is **not consistent** if $(1/2 - p_n)^2 = o(\sqrt{\log n / n})$ and

$$\beta_n = o\left(\frac{\log n}{(1/2 - p_n)^2 n}\right).$$

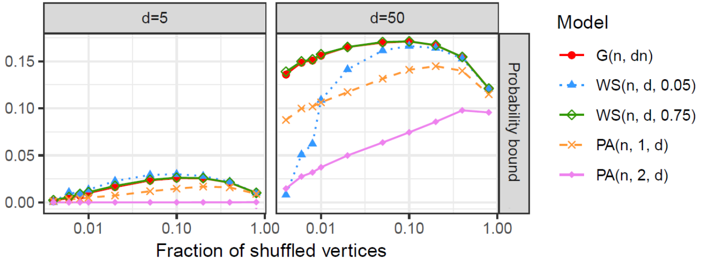
- $\widehat{Q}_{\mathrm{MLE}}$ is **consistent** if $d_n = o(\beta_n^2 n)$ and

$$\beta_n \geq c\sqrt{\frac{\log n}{n\,(1/2 - p_n)^2}}.$$

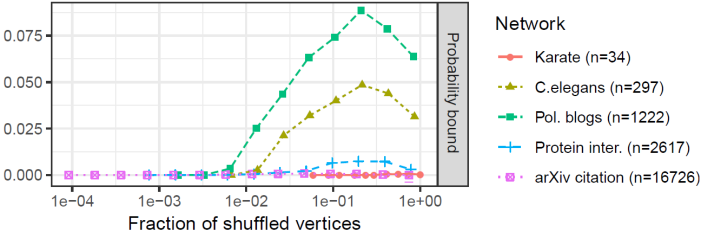
Matchability on random graphs

Measure matching feasibility: Upper bound for the noise probability tolerated by a graph based on the theory

- Erdös-Rényi $G(n, dn)$
- Watts-Strogatz small-world $WS(n, d, \beta)$
- Preferential attachment $PA(n, \gamma, d)$



Matchability on real networks



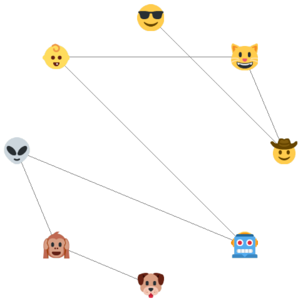
Outline

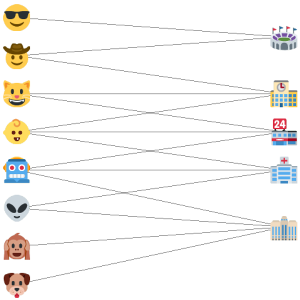
Graph matching in errorfully observed networks

Graph matching between bipartite and unipartite networks

Data integration

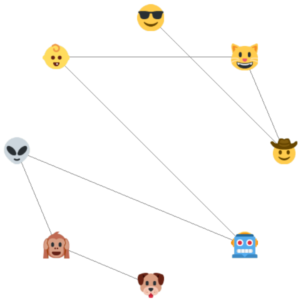
- Data are often collected from different sources or modalities
- In particular, now consider unipartite and bipartite graphs

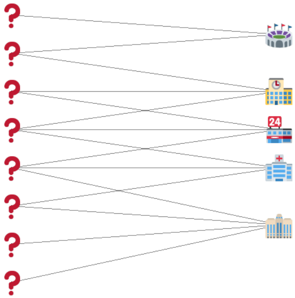




Data integration

- Data are often collected from different sources or modalities
- In particular, now consider unipartite graphs and bipartite data

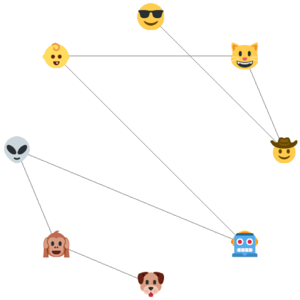


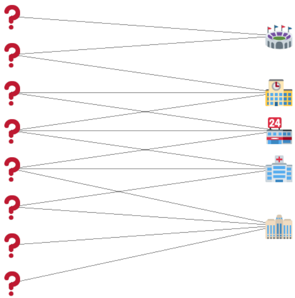


Graph matching

Goal: graph matching between bipartite and unipartite networks

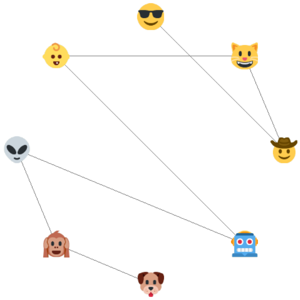
- *Methodology*: joint model based on undirected graphical models
- *Graph matching*: use MLE to find unshuffling permutation
- *Optimization*: non-convex relaxation via graphical lasso and fast QAP

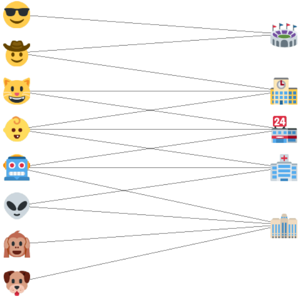




Graph matching formulation

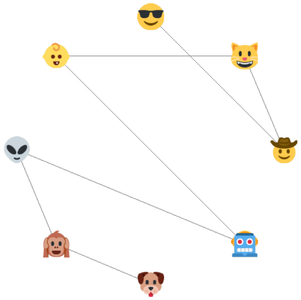
$A \in \{0,1\}^{n \times n}$ adjacency matrix, $B \in \mathbb{R}^{n \times m}$ incidence or data matrix

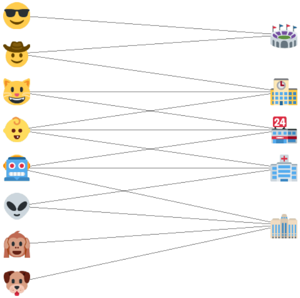




Graph matching formulation

$A \in \{0,1\}^{n \times n}$ adjacency matrix, $B \in \mathbb{R}^{n \times m}$ incidence or data matrix





Undirected graphical model for B conditioned on A

- Local Markov property: edges of vertex i are conditionally independent to other edges given the values of the neighbors of i. If X is a column of B

$$X_i \perp X_{[n] \setminus \mathcal{N}_i(A) \cup \{i\}} \mid X_{\mathcal{N}_i(A)}, \qquad \forall i \in [n].$$

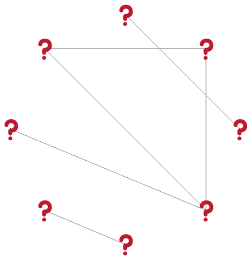
Graph matching formulation

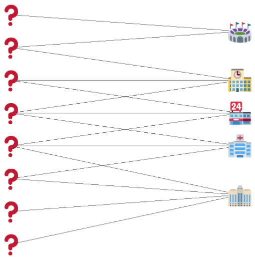
- Generalized linear model distributions (Yang et al., 2012) to make the problem tractable.

$$f_X\left(x_i \mid x_{[n]\setminus\{i\}}\right) \propto \exp\left(\beta_i x_i + \sum_{j \in \mathcal{N}_i(W)} \Theta_{ij} x_i x_j - 2\Theta_{ii} C(x_i)\right),$$

  - $\Theta_{ij} = 0$ if $A_{ij} = 0$ (local Markov property)
  - Special cases: Ising model, Gaussian graphical model

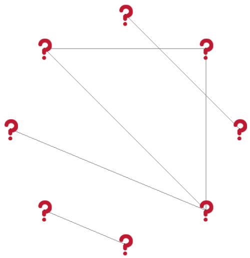
Bipartite-to-unipartite graph matching formulation

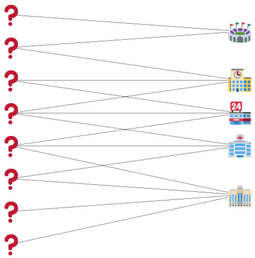




- Graph matching: we observe $A' = P^* A (P^*)^T$ for a permutation P^*.

Bipartite-to-unipartite graph matching formulation





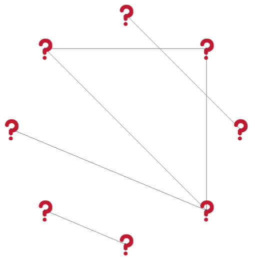
- Graph matching: we observe $A' = P^* A (P^*)^T$ for a permutation P^*.
- Solve restricted *maximum likelihood estimation*:

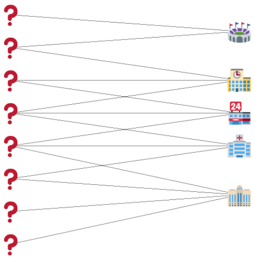
$$\left(\widehat{P}, \widehat{\Theta}\right) = \underset{P,\Theta}{\operatorname{argmax}} \quad \ell(\Theta)$$

$$\text{subject to} \quad \Theta_{ij}(1 - (P^T A' P)_{ij}) = 0, \quad i \neq j,$$

P is a permutation matrix,

Exact graph matching recovery





Theorem (A., Priebe, Lyzinski, 2021)

Suppose that $\Theta^*_{ij} \neq 0$ *if* $((P^*)^\top A P^*)_{ij} = 1, i \neq j$.

Under **regularity conditions**, *if*

$$\underbrace{\min_{P \neq I} \|A - P^\top A P\|_F^2}_{\text{Graph matching difficulty}} \geq \underbrace{C \frac{(\|A\|_F^2 + n) \log n}{m}}_{\text{Graphical model estimation error}},$$

Graph matching difficulty (Lyzinski et al., 2016; A. et al., 2021)

Graphical model estimation error (Rothman et al., 2008)

then $\widehat{P} = P^*$ *with high probability.*

Graph matching algorithm

- Maximum likelihood estimation is NP-hard!

Graph matching algorithm

- Maximum likelihood estimation is NP-hard!
- Strategy to find an approximate solution:
    1. Relax permutation Q to a doubly stochastic matrix D
    2. Write the problem in a Lagrangian formulation
    3. Alternating optimization for D and Θ.

Graph matching algorithm

- Maximum likelihood estimation is NP-hard!
- Strategy to find an approximate solution:
  1. Relax permutation Q to a doubly stochastic matrix D
  2. Write the problem in a Lagrangian formulation
  3. Alternating optimization for D and Θ.
- All steps have efficient solutions!

- For Gaussian graphical models, the new optimization problem is

$$\underset{D,\Theta}{\operatorname{argmax}}\left\{ \log\det\Theta - \mathsf{trace}(\hat{\Sigma}\Theta) - \lambda \sum_{i\neq j} \left|(1-(D^T A D)_{ij})\Theta_{ij}\right| \right\}$$

  1. Optimization for Θ: weighted graphical lasso (Friedman et al. 2008)
  2. Optimization for D: quadratic assignment problem (Vogelstein et al., 2014, Lyzinski et al., 2016)

Matching via inverse covariance estimation

Algorithm 1 Unipartite to bipartite matching via penalized inverse covariance estimation

Input: Adjacency matrix A, incidence matrix B.

for each $\lambda \in \{\lambda_s\}_{s=1}^{S^*}$ **do**

Initialize $\hat{D}^{(1,\lambda)} = \frac{1}{n} 1_n 1_n^\top$.

for $t = 1, \ldots, T^*$, or until convergence **do**

Update $\hat{\Theta}^{(t,\lambda)}$ by solving (3.6).

Update $\hat{D}^{(t+1,\lambda)}$ by solving (3.7).

Set $\hat{P}^{(t+1,\lambda)}$ as the projection of $\hat{D}^{(t,\lambda)}$ into Π_n.

end for

end for

Choose the permutation with the largest value of $\hat{\ell}(\hat{\Theta}_P)$ among the permutations $P \in \{P^{(t,\lambda_s)}, s \in [S^*], t \in [T^*]\}$.

Output: Permutation $\widehat{P}$, inverse covariance estimate $\hat{\Theta}_{\widehat{P}}$.

Matching via penalized pseudolikelihood

- Use pseudolikelihood when likelihood is intractable (e.g., Ising model).

Algorithm 2 Unipartite to bipartite matching via penalized pseudolikelihood

---

Input: Adjacency matrix A, incidence matrix B.

  for each $\lambda \in \{\lambda_s\}_{s=1}^{S^*}$ **do**

    Initialize $\hat{D}^{(1,\lambda)} = \frac{1}{n} 1_n 1_n^\top$.

    for $t = 1, \ldots, T^*$, or until convergence **do**

      for $j = 1, \ldots, n$ **do**

        Update $(\hat{\Theta}_j^{(t,\lambda)}, \hat{\beta}_j^{(t,\lambda)})$ by solving (3.9).

      end for

      Update $\hat{D}^{(t+1,\lambda)}$ by solving (3.7).

      Set $\hat{P}^{(t+1,\lambda)}$ as the projection of $\hat{D}^{(t,\lambda)}$ into Π_n.

    end for

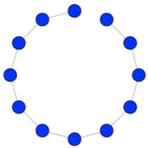
  end for

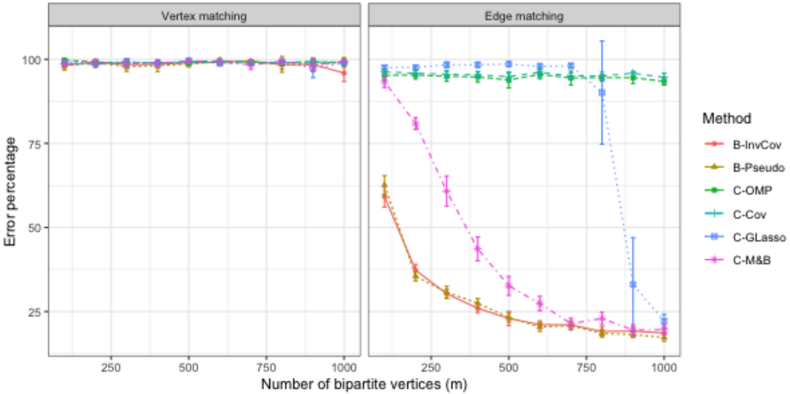
  Choose the permutation with the largest value of $\widetilde{\ell}(\hat{\Theta}_P)$ among $P \in \{P^{(t,\lambda_s)}, s \in [S^*], t \in [T^*]\}$.

Output: Permutation $\widehat{P}$, estimated parameters $\hat{\Theta}_{\widehat{P}}$ and $\hat{\beta}_{\widehat{P}}$.

Simulation experiment 1

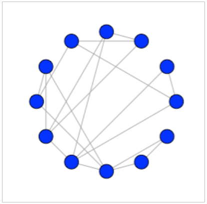
- A (unipartite) is a **chain graph**, B follows Ising model
- Graphical model estimation: easy
- Graph matching: hard

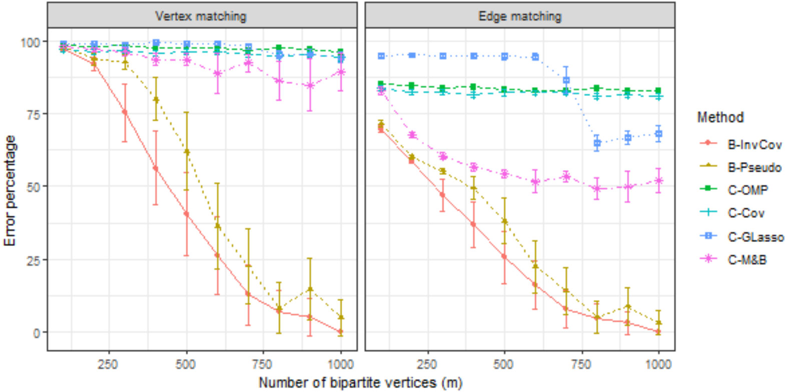




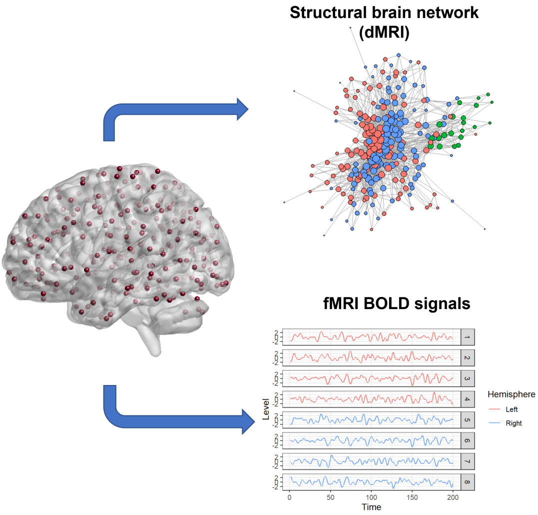
Simulation experiment 2

- A (unipartite) is an **Erdős-Rényi graph**
- Graphical model estimation: hard
- Graph matching: easy



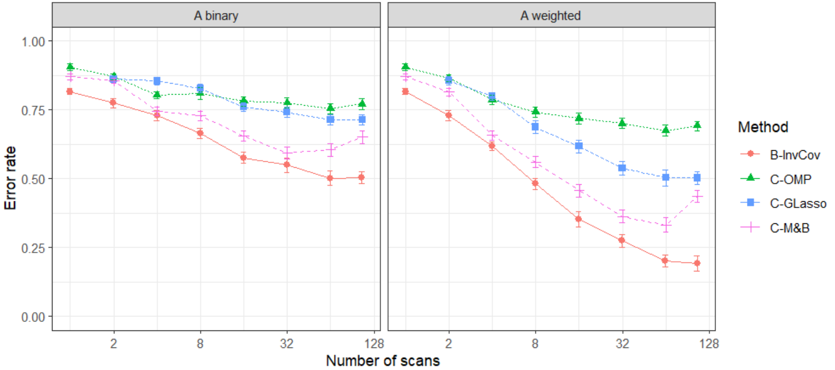


Magnetic resonance imaging (MRI) data



Structural and functional MRI data (Zuo et al, 2014).

MRI data



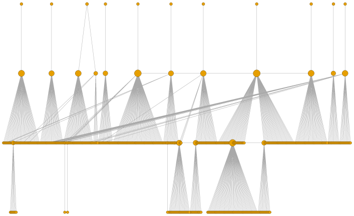
A binary
A weighted
Error rate
1.00
0.75
0.50
0.25
0.00
2
8
32
128
Number of scans
Method
B-InvCov
C-OMP
C-GLasso
C-M&B

Concluding remarks

- Integrating multiple data sources often requires to match the units
- Statistical approaches based on random graph models for matching
- Combining information may improve performance
- Graph matching with other data structures? Networks with attributes, multilayer or time-varying graphs.
- Statistical inference for multiple networks? (after matching)

Future directions

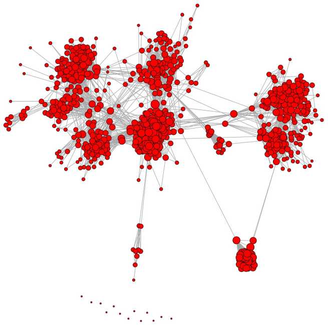
- Work in progress: academic and collaboration networks (data collected by Xingyu Liu and Yufan Li)
- Integration of different network data sources:
    - Efficient graph matching methods
    - Joint statistical models for heterogeneous modalities
    - Statistical inference problems

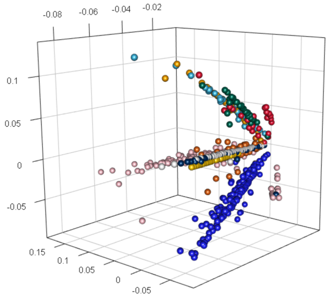


Advertisement

New course for Spring 2024!

- Special Topics in Network Data Analysis (STAT 689)
- Tue - Thu 11:10 - 12:25 (3 credits)
- Supported by TAMIDS Course Development Program





Thank you!

Questions?

✉jarroyo@tamu.edu
https://jesus-arroyo.github.io/

Main references:

- Arroyo, J., Sussman, D.L., Priebe, C.E. and Lyzinski, V. (2021) “Maximum Likelihood Estimation and Graph Matching in Errorfully Observed Networks”, *Journal of Computational and Graphical Statistics* 30:4.
- Arroyo, J., Priebe, C.E. and Lyzinski, V. (2021) “Graph matching between bipartite and unipartite networks: to collapse, or not to collapse, that is the question”, *IEEE Trans. on Network Science and Engineering*