Sparse Network Estimation

Olga Klopp

ESSEC BUSINESS SCHOOL

CREST
CENTER FOR RESEARCH
IN ECONOMICS AND STATISTICS

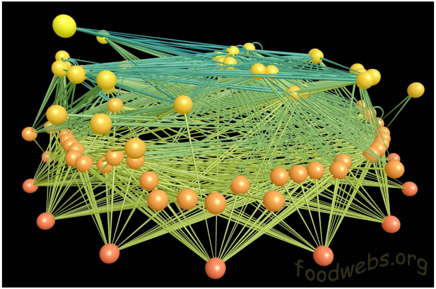
Joint works with

Alexandre Tsybakov

Nicolas Verzelen

Solenne Gaucher

Network model



East-river trophic network [Yoon et al.(04)]

Approach

- Model-based statistical analysis.
- The modeling of real networks as random graphs.

Stochastic Block-Model (SBM) Holland et al. (1980)

- Fit observed networks to parametric or non-parametric models of random graphs.
- **SBM** popular in applications: it allows to generate graphs with a community structure
- Parameters:
  - Partition of n nodes into k disjoint groups $\{C_1, \ldots, C_k\}$
    - each node i is associated with a community $z(i)$
    - $z : [n] \to [k]$: the index function
    - z: a parameter to estimate (the conditional SBM), or a latent variable
  - Symmetric $k \times k$ matrix Q of inter-community edge probabilities.
    - Any two vertices $u \in C_i$ and $v \in C_j$ are connected with probability Q_{ij}
- **Regularity Lemma:** basic approximation units for more complex models.

Inhomogeneous random graph model

- We observe the $n \times n$ adjacency matrix $\mathbf{A} = (\mathbf{A}_{ij})$ of a graph
- A_{ij} are Bernoulli random variables with parameter Θ_{ij}
- Θ_0 is the $n \times n$ symmetric matrix with entries (Θ_{ij}) (the matrix of probabilities associated to the graph)
  - vertices i and j are connected by an edge with probability $\boldsymbol{\Theta}_{ij}$ independently from any other edge
  - sparsity parameter $\rho_n = \max_{ij} \boldsymbol{\Theta}_{ij} \to 0$ and $\rho_n \geq 1/n$
- Given a single observation $\mathbf{A}$, we want to estimate $\boldsymbol{\Theta}_0$

Minimax rate for sparse SBM in Frobenius norm

The best rate of convergence that any estimator may achieve:

K., Tsybakov & Verzelen (2017)

$$\inf_{\widehat{\mathbf{\Theta}}} \sup_{\mathbf{\Theta}_0 \in \mathcal{T}[k,\rho_n]} \mathbb{E}_{\mathbf{\Theta}_0}\left[\frac{1}{n^2}\left\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}_0\right\|_2^2\right] \asymp \min\left\{\rho_n\left(\frac{\log k}{n} + \frac{k^2}{n^2}\right), \rho_n^2\right\}$$

- $\rho_n = 1$: **Gao et al.(2014)**, the minimax rate over $\mathcal{T}[k,1]$

$$\frac{k^2}{n^2} + \frac{\log k}{n}$$

  - $k > \sqrt{n\log(k)}$: nonparametric rate $\frac{k^2}{n^2}$
  - $k < \sqrt{n\log(k)}$: clustering rate $\frac{\log k}{n}$

Sparse network estimation problem

- The optimal rates can be achieved by the **Least Squares Estimator**
- But: not realizable in polynomial time
- Better choices:
  - Maximum Likelihood Estimator
  - Hard thresholding estimator
  - ...

Maximum Likelihood Estimator

- **Wolfe and Olhede (2013)**, **Bickel et al (2013)**, **Amini et al (2013)**, **Celisse et al (2012)** , **Tabouy et al (2017)** ...
- Also NP hard ...
- Computationally efficient approximations:
  - Pseudo-likelihood methods
  - Variational approximation
- Quite successful in practice

Is MLE minimax optimal?

Convergence rate for the MLE

The conditional log-likelihood:

$$\mathcal{L}(\mathbf{A}; z, \mathbf{Q}) = \sum_{i<j} \mathbf{A}_{ij} \log(\mathbf{Q}_{z(i)z(j)}) + (1 - \mathbf{A}_{ij}) \log(1 - \mathbf{Q}_{z(i)z(j)})$$

The maximum log-likelihood estimator of $\mathbf{\Theta}^*$:

$$(\widehat{\mathbf{Q}}, \widehat{z}) \in \underset{Q \in [0,1]^{k \times k}, z \in \mathcal{Z}_{n,k}}{\operatorname{argmax}} \mathcal{L}(\mathbf{A}; z, \mathbf{Q}).$$

$\mathcal{Z}_{n,k}$ the set of all possible mappings z from $[n]$ to $[k]$

Theorem (Gaucher & K., 2021)

With high probability

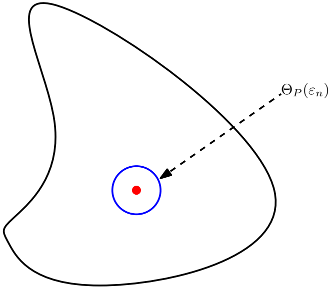
$$\frac{1}{n^2} \|\mathbf{\Theta}_0 - \widehat{\mathbf{\Theta}}_{ML}\|_2^2 \leq C_{\rho_n, \gamma_n} \, \rho_n \Big(\frac{\log k}{n} + \frac{k^2}{n^2} \Big).$$

- $0 < \gamma_n \leq \mathbf{\Theta}_{ij} \leq \rho_n < 1$
- **Minimax optimal** if $\gamma_n \asymp \rho_n$

Variational approximation

- The optimization of the likelihood function requires a search over the set of k^n labels $\Rightarrow$ MLE is computationally intractable
- Solution: **Variational approximation**
  - serves to approximate the posterior
    - distributions for the unobserved variables (parameters, latent variables)
    - often hard-to-solve integrals
  - Kullback–Leibler divergency as a measure of good approximation
  - Assuming the unknown variables can be partitioned so that each partition is independent: **the mean-field approximation**
  - Often results in easy to compute interactive algorithms

Subhabrata's talk: variational approximation can lead to a quite accurate approximation



S. Sen's courtesy

The mean field approximation works exceptionally well for the SBM

Variational approximation to the MLE

- Celisse et al (2008) and Bickel et al (2013): variational approximation
  - asymptotic normality for variational methods for parameter estimates of stochastic block data
- The problem of community detection: Edoardo et al (2008), Hofman et al (2008), Zhang et al (2020), Razaee et al (2019) . . .
- Gaucher & K. (2021):
  - optimal statistical accuracy
  - labels recovery

SBM with random labels

- Nodes are classified into k communities:
  - each node i is associated with a community $z(i)$
  - $z : [n] \to [k]$: the index function
  - z: a parameter to estimate (the conditional SBM), or a latent variable
- **The indexes follow a multinomial distribution:**

$$\forall i \quad z(i) \overset{i.i.d}{\sim} \mathcal{M}(1; \alpha)$$

  - $\forall l \in [k]$, α_l is the probability that node i belongs to the community l
  - $\alpha_k n$ is the expected size of community k
  - the probabilities of connection are given by a $k \times k$ matrix $\mathbf{Q}$
- We consider a SBM with parameters $(\alpha, \mathbf{Q})$.

Variational approximation to the MLE

- SBM with parameters $(\alpha, \mathbf{Q})$
- The likelihood of the observed adjacency matrix $\mathbf{A}$:

$$\mathfrak{l}(\mathbf{A}; \alpha, \mathbf{Q}) \quad = \quad \sum_{z \in \mathcal{Z}_{n,k}} \left(\prod_{i \leq n} \alpha_{z(i)} \right) \exp\left(\mathcal{L}(\mathbf{A}; z, \mathbf{Q})\right).$$

- the maximization still requires to evaluate the expectation of the label function z by summing over k^n possible labels
- Solution: use the **mean-field approximation**

Mean-field approximation

$$l(\mathbf{A};\alpha,\mathbf{Q}) \quad = \quad \sum_{z\in\mathcal{Z}_{n,k}} \left(\prod_{i\leq n}\alpha_{z(i)}\right)\exp\left(\mathcal{L}(\mathbf{A};z,\mathbf{Q})\right).$$

- Approximate the posterior distribution of z by a simpler distribution:
  - the posterior distribution $\mathbb{P}\left(\cdot|\mathbf{A},\alpha,\mathbf{Q}\right)$ is approximated by a multinomial distribution $\mathbb{P}_\tau$, s.t. $\mathbb{P}_\tau(z)=\prod_{1\leq i\leq n}\mathcal{M}(z|\tau^i)$
  - $\tau^i=\left(\tau_1^i,...,\tau_K^i\right)$
  - $\tau=\left(\tau^1,...,\tau^n\right)$
- Use the KL-divergence as a measure of how well our approximation fits the true posterior

Variational approximation to the MLE

The variational estimator:

$$\left(\widehat{\alpha}^{VAR}, \widehat{\mathbf{Q}}^{VAR}, \widehat{\tau}^{VAR}\right) = \underset{\alpha\in\mathcal{A},\mathbf{Q}\in\mathcal{Q},\tau\in\mathcal{T}}{\operatorname{argmax}} \mathcal{J}(\mathbf{A};\tau,\alpha,\mathbf{Q}) \quad (1)$$

$$\text{for} \quad \mathcal{J}(\mathbf{A};\tau,\alpha,\mathbf{Q}) = \mathfrak{l}(\mathbf{A};\alpha,\mathbf{Q}) - KL\left(\mathbb{P}_\tau(\cdot)||\mathbb{P}\left(\cdot|\mathbf{A},\alpha,\mathbf{Q}\right)\right)$$

- $\mathcal{A}$, $\mathcal{Q}$ and $\mathcal{T}$: the parameter spaces for the parameters α, $\mathbf{Q}$ and τ
- KL: the Kullback-Leibler divergence
- $\mathcal{J}(\mathbf{A};\tau,\alpha,\mathbf{Q})$ provides a lower bound on $\mathfrak{l}(\mathbf{A};\alpha,\mathbf{Q})$

EM algorithm

The expectation - maximization (EM) algorithm Tabouy et al (2020):

- Estimation Step: given parameters $(\alpha, \mathbf{Q})$, the variational parameter τ maximizing $\mathcal{J}(\mathbf{A}; \tau, \alpha, \mathbf{Q})$ is given by the fixed point equation :

$$\tau_k^i \propto \alpha_k \prod_{j \neq i} \prod_{l \leq K} \left(\mathbf{Q}_{kl}^{\mathbf{A}_{ij}} \left(1 - \mathbf{Q}_{kl}\right)^{1-\mathbf{A}_{ij}} \right)^{\tau_l^j} ;$$

- Maximisation Step: given parameter τ, the parameters $(\alpha, \mathbf{Q})$ maximizing $\mathcal{J}(\mathbf{A}; \tau, \alpha, \mathbf{Q})$ are given by

$$\alpha_k = \frac{\sum_i \tau_k^i}{n} \, , \, \mathbf{Q}_{kl} = \frac{\sum_{i \neq j} \tau_k^i \tau_l^j \mathbf{A}_{ij}}{\sum_{i \neq j} \tau_k^i \tau_l^j}.$$

Statistical guarantees for the variational estimator

- Celisse et al (2008) and Bickel et al (2013):
  - maximizing $\max_{\tau \in \mathcal{T}} \mathcal{J}(\mathbf{A}; \tau, \alpha, \mathbf{Q})$ is equivalent to maximising $\mathfrak{l}(\mathbf{A}; \alpha, \mathbf{Q})$
  - the estimator obtained by maximizing $\mathfrak{l}(\mathbf{A}; \alpha, \mathbf{Q})$ converges to the true parameters $(\alpha, \mathbf{Q})$
  - $(\widehat{\alpha}^{VAR}, \widehat{\mathbf{Q}}^{VAR})$ also converges to $(\alpha, \mathbf{Q})$
  - does not provide guarantees on the recovery of the true labels z

The label estimator

- The label estimator $\widehat{z}^{VAR}$:

$$\forall\ i \leq n,\ \ \widehat{z}^{VAR}(i) \triangleq \underset{k\leq K}{\operatorname{argmax}} \left(\widehat{\tau}^{VAR}\right)_k^i$$

- Replace $\widehat{\mathbf{Q}}^{VAR}$ by the empirical mean estimator:

$$\widehat{\mathbf{Q}}_{ab}^{ML-VAR} \triangleq \frac{\sum\limits_{i\in(\widehat{z}^{VAR})^{-1}(a),j\in(\widehat{z}^{VAR})^{-1}(b),i\neq j} \mathbf{A}_{ij}}{n_{ab}}$$

-

$$n_{ab}(\widehat{z}^{VAR}) = \begin{cases} |(\widehat{z}^{VAR})^{-1}(a)| \times |(\widehat{z}^{VAR})^{-1}(b)| & \text{if } a \neq b \\ |(\widehat{z}^{VAR})^{-1}(a)| \times \left(|(\widehat{z}^{VAR})^{-1}(a)| - 1\right) & \text{otherwise} \end{cases}$$

- Define $\widehat{\boldsymbol{\Theta}}^{VAR}$ as

$$\widehat{\boldsymbol{\Theta}}_{i\neq j}^{VAR} = \widehat{\mathbf{Q}}_{\widehat{z}^{VAR}(i),\widehat{z}^{VAR}(j)}^{ML-VAR},\ \ \widehat{\boldsymbol{\Theta}}_{ii}^{VAR} = 0.$$

Statistical guarantees for the variational estimator

This new estimator $\left(\widehat{z}^{VAR}, \widehat{\mathbf{Q}}^{ML-VAR}\right)$ is minimax optimal:

Theorem (Gaucher & K., 2021)

Assume that $\mathbf{Q}^0$ has no identical columns and the sparsity inducing sequence ρ_n satisfies $\rho_n \gg \log(n)/n$. Then, there exists a constant $C_{\mathbf{Q}^0} > 0$ depending on $\mathbf{Q}^0$ such that

$$\mathbb{P}\left(\left\|\mathbf{\Theta}_0 - \widehat{\mathbf{\Theta}}^{VAR}\right\|_2^2 \leq C_{\mathbf{Q}^0}\rho_n\left(k^2 + n\log(k)\right)\right) \underset{n\to\infty}{\to} 1$$

1. $\alpha = \alpha^0$ for some fixed α^0 such that $\alpha^0_a > 0$ for any $a \in \{1, ..., k\}$
2. $\mathbf{Q} = \rho_n \mathbf{Q}^0$ for some fixed $\mathbf{Q}^0 \in (0,1)^{k\times k}$ such that $\sum_{a,b=1}^{k} \alpha^0_a \alpha^0_b \mathbf{Q}^0_{ab} = 1$

How does it work ?

- Variational approximation to the MLE has been used for estimation of $(\mathbf{Q}, \alpha)$
- We show that **both the maximum likelihood estimator and its variational counterpart can perfectly recover all labels**:
  - with large probability, there exists a permutation σ of $\{1, ..., K\}$ such that $\left(\widehat{z}^{VAR}(\sigma(k))\right)_{k \leq K} = \left(\widehat{z}(k)\right)_{k \leq K}$
  - (under certain conditions) MLE recovers the true labels
- exact recovery of the labels have already been established in this regime under more restricted assumptions (see **Abbe (2018)**):
  - the SBM is symmetric, assortative and has balanced communities

Non-parametric Model

- SBM does not allow to analyze the fine structure of extremely large networks, in particular when the number of groups is growing.
- Non-parametric models of random graphs: **Graphon Model**
  - Graphons are symmetric measurable functions

$$W : [0,1]^2 \to [0,1].$$

  - Play a central role in the recent theory of graphs limits: every graph limit can be represented by a graphon.
  - Graphons give a natural way of generating random graphs.

Graphon Model

- **Graphon Model:**
  - $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ are latent i.i.d. uniformly distributed on $[0, 1]$.

$$\boldsymbol{\Theta}_{ij} = W_0(\xi_i, \xi_j).$$

  - The diagonal entries $\boldsymbol{\Theta}_{ii}$ are zero and $\boldsymbol{\Theta}_0 = (\boldsymbol{\Theta}_{ij})$
  - Given $\boldsymbol{\Theta}_0$ the graph is sampled according to the **inhomogeneous random graph model**:
    - vertices i and j are connected by an edge with probability $\boldsymbol{\Theta}_{ij}$ independently from any other edge.
  - If W_0 is a step-function with k steps, the graph is distributed as a SBM with k groups.

Sparse Graphon Model

- The expected number of edges $\asymp n^2 \Rightarrow$ **dense** case.
- In real life networks often **sparse**
- **Sparse Graphon Model:**
  - Take $\rho_n > 0$ such that $\rho_n \to 0$ as $n \to \infty$.
  - The adjacency matrix $\mathbf{A}$ is sampled according to graphon W_0 with scaling parameter ρ_n:

$$\mathbf{\Theta}_{ij} = \rho_n W_0(\xi_i, \xi_j),\ i < j.$$

  - $\rho_n =$ "expected proportion of non-zero edges",
  - the number of edges is of the order $O(\rho_n n^2)$,
    - $\rho_n = 1$ dense case
    - $\rho_n = 1/n$ very sparse

Graphon: invariance with respect to the change of labeling

- Graphon estimation is **more challenging** than probability matrix estimation
- Multiple graphons can lead to the same distribution on the space of graphs of size n.
- The topology of a network is **invariant with respect to any change of labeling** of its nodes
- We consider **equivalence classes** of graphons defining the same probability distribution on random graphs.

Loss function for graphon estimation

- Consider a sparse graphon $f(x, y) = \rho_n W(x, y)$
- $\tilde{f}(x, y)$ estimator of $f(x, y)$
- The squared error is defined by

$$\delta^2(f, \tilde{f}) := \inf_{\tau \in \mathcal{M}} \int\int_{(0,1)^2} |f(\tau(x), \tau(y)) - \tilde{f}(x, y)|^2 \mathrm{d}x\mathrm{d}y$$

$\mathcal{M}$ is the set of all measure-preserving bijections $\tau : [0, 1] \to [0, 1]$

Property (Lovász 2012)

$\delta(\cdot, \cdot)$ defines a metric on the quotient space $\mathcal{W}$ of graphons.

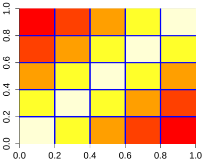
From probability matrix estimation to graphon estimation

- To any $n \times n$ probability matrix $\boldsymbol{\Theta}$ we can associate a graphon.

- Given a $n \times n$ matrix $\boldsymbol{\Theta}$ with entries in $[0, 1]$, define the **empirical graphon** $\widetilde{f}_{\boldsymbol{\Theta}}$ as the following piecewise constant function:

$$\widetilde{f}_{\boldsymbol{\Theta}}(x, y) = \boldsymbol{\Theta}_{\lceil nx \rceil, \lceil ny \rceil}$$

for all x and y in $(0, 1]$.



- This provides a way of deriving an estimator of the graphon function $f(\cdot, \cdot) = \rho_n W(\cdot, \cdot)$ from **any** estimator of the probability matrix $\boldsymbol{\Theta}_0$.

From probability matrix estimation to graphon estimation

- **Empirical graphon** $\widetilde{f}_{\mathbf{\Theta}}(x,y) = \mathbf{\Theta}_{\lceil nx \rceil, \lceil ny \rceil}$.
- For any estimator $\widehat{\mathbf{T}}$ of $\mathbf{\Theta}_0$:

$$\mathbb{E}\left[\delta^2(\widetilde{f}_{\widehat{\mathbf{T}}}, f)\right] \le 2\mathbb{E}\left[\frac{1}{n^2}\|\widehat{\mathbf{T}} - \mathbf{\Theta}_0\|_F^2\right] + 2\underbrace{\mathbb{E}\left[\delta^2\left(\widetilde{f}_{\mathbf{\Theta}_0}, f\right)\right]}_{\textbf{agnostic error}}$$

(from the triangle inequality). Here, $\widetilde{f}_{\widehat{\mathbf{T}}}$ and $\widetilde{f}_{\mathbf{\Theta}_0}$ are empirical graphons.

Bound for the δ-risk of step-function graphon

Step function graphons: For some $k \times k$ symmetric matrix $\mathbf{Q}$ and some $\phi : [0,1] \to [k]$,

$$W(x,y) = \mathbf{Q}_{\phi(x),\phi(y)} \quad \text{for all } x, y \in [0,1] \; .$$

Theorem (K., Tsybakov and Verzelen, 2017)

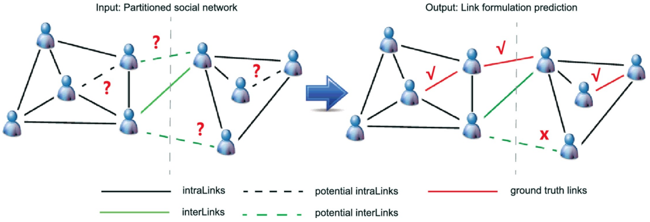
Consider the ρ_n-sparse step-function graphon model $\mathcal{W}[k,\rho_n]$.

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{W}[k,\rho_n]} \mathbb{E}\left[\delta^2\left(\widehat{f}, f\right)\right] \asymp \left[\rho_n\left(\frac{k^2}{n^2} + \frac{\log(k)}{n}\right) + \boldsymbol{\rho_n^2}\sqrt{\frac{\mathbf{k}}{\mathbf{n}}}\right].$$

Missing Links

Real-life networks are only partially observed

- Exhaustive exploration of all interactions in a network is expensive
- Survey data: non-response or drop-out of participants
- Online social network data: sub-sample of the network



A balanced modularity maximization link prediction model in social networks [Wu et al.(2017)]

Conditional maximum likelihood estimator

- the log-likelihood function with respect to the observed entries of the adjacency matrix $\mathbf{A}$ and sampling matrix $\mathbf{X}$:

$$\begin{aligned}\mathcal{L}_{\mathbf{X}}(\mathbf{A}; z, \mathbf{Q}) = \sum_{1 \le i < j \le n} & \mathbf{X}_{ij} \big(\mathbf{A}_{ij} \log(\mathbf{Q}_{z(i)z(j)}) \\ & + (1 - \mathbf{A}_{ij}) \log(1 - \mathbf{Q}_{z(i)z(j)}) \big) \\ = & \sum_{a \le b} \log(\mathbf{Q}_{ab}) \sum_{\substack{i \in z^{-1}(a),\ j \in z^{-1}(b) \\ i \ne j}} \mathbf{X}_{ij} \mathbf{A}_{ij} \\ & + \sum_{a \le b} \log(1 - \mathbf{Q}_{ab}) \sum_{\substack{i \in z^{-1}(a),\ j \in z^{-1}(b) \\ i \ne j}} \mathbf{X}_{ij} (1 - \mathbf{A}_{ij})\end{aligned}$$

- X_{ij} are iid Bernoulli (p)

Theorem (Gaucher & K., 2021)

Assume that $\mathbf{Q}^0$ *has no identical columns and the sparsity inducing sequence* ρ_n *satisfies* $\rho_n \gg \log(n)/(pn)$. *Then,*

$$\mathbb{P}\left(\widehat{z}^{VAR} \sim \widehat{z}\right) \to 1$$

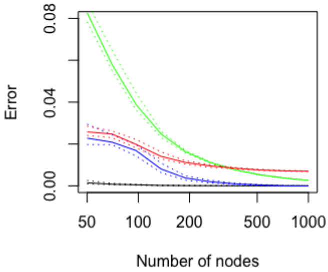
when $n \to \infty$. *Moreover, there exists a constant* $C_{\mathbf{Q}^0} > 0$ *depending on* $\mathbf{Q}^0$ *such that*

$$\mathbb{P}\left(\left\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}^{VAR}\right\|_2^2 \leq \frac{C_{\mathbf{Q}^0}\left(k^2 + n\log(k)\right)}{p}\right) \underset{n\to\infty}{\to} 1.$$

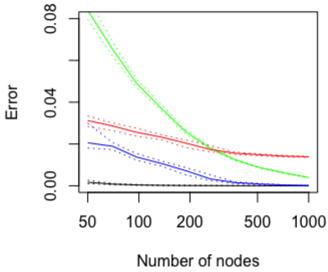
Empirical performances of the variational approximation of MLE

We compare the variational approximation to the MLE to

- missSBM
- softImpute
- the oracle estimator with knowledge of the label z^*



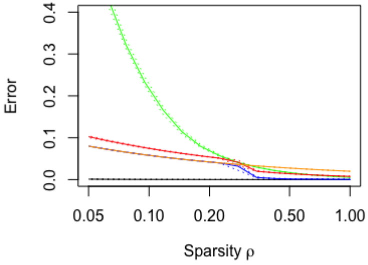
Assortative SBM



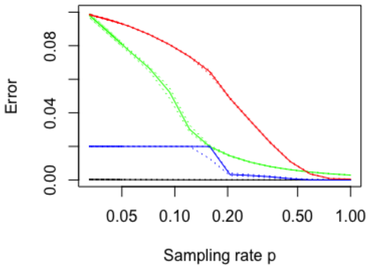
Mixed SBM

Robustness against sparsity and missing observations

Error of connection probabilities estimation as a function of the sparsity parameter ρ and of the sampling rate p ($n = 500$):



Robustness against sparsity



Robustness against missing observations.

Interactions within a elementary school

- Network of interactions within a French elementary school Stehle et al (2011) :
  - Physical interactions between 222 children divided into 10 classes and their 10 teachers
  - Two consecutive days
  - Homogeneous degrees (the maximum degree is 41, the minimum degree is 5 and the mean degree is 20)
  - Strong community structure. Therefore, we expect the networks of interactions to be well approximated by a stochastic block model
- Two outcomes of the same random network model:
  - We use the observations collected on Day 1 to estimate the matrix $\mathbf{\Theta}^*$
  - Evaluate the estimators on the network of Day 2

Interactions within a elementary school

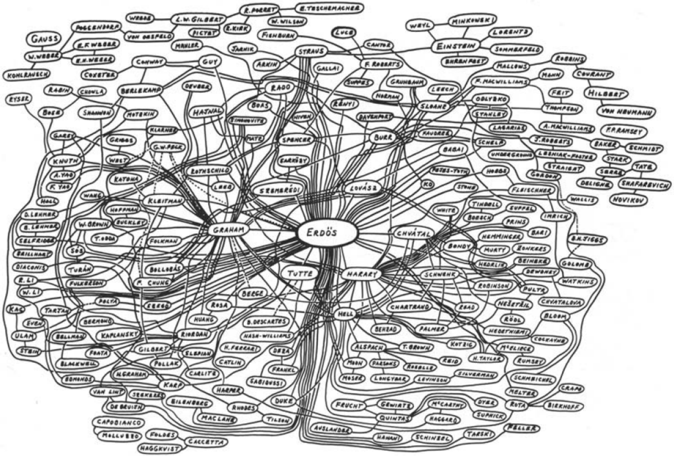
Estimator	$\widehat{\mathbf{\Theta}}^{VAR}$	$\widehat{\mathbf{\Theta}}^{missSBM}$	$\widehat{\mathbf{\Theta}}^{SVT}$	$\widehat{\mathbf{\Theta}}^{naive}$
$\|\mathbf{X} \odot (\mathbf{A} - \widehat{\mathbf{\Theta}})\|_2^2 / \|\mathbf{X} \odot \mathbf{A}\|_2^2$	0.312	0.317	0.357	0.541

Table: Normalized squared distance between the observed adjacency matrix for the network on interaction on Day 2, and its predicted value.

- The naive estimator $\widehat{\mathbf{\Theta}}^{naive}$:
  - $\widehat{\mathbf{\Theta}}_{ij}^{naive} = 1$ if an interaction between i and j has been recorded on Day 1
  - $\widehat{\mathbf{\Theta}}^{naive} = 0$ if no such interaction has been recorded
  - $\widehat{\mathbf{\Theta}}_{ij}^{naive} = d/n$ if the information is missing, where d is the average degree of the graph for Day 1.

Conclusion

- **Least Squares Estimator**:
  - attains the optimal rates in a minimax sense
  - not realizable in polynomial time
- **(variational) MLE:**
  - minimax optimal
  - allows labels recovery
- Variational MLE has good performances in practice
- Can be used for **Link Prediction**


GAUSS
W.WEBER
CONWAY
GUY
BERLEKAMP
STRAUS
EINSTEIN
MINKOWSKI
LORENTZ
SOMMERFELD
WEYL
CANTOR
HILBERT
VON NEUMANN
COURANT
ROBBINS
RADO
RÉNYI
SLOANE
HAJNAL
SPENCER
BURR
BABAI
KNUTH
KATONA
KLEITMAN
SZEMERÉDI
LOVÁSZ
GRAHAM
ERDÖS
CHVÁTAL
TUTTE
HARARY
BONDY
F. CHUNG
BERGE
ROSA
HELL
CHARTRAND
PALMER
ALSPACH
KARP
DUKE
FRUCHT
QUINTAS
ROTA
BIRKHOFF
TARSKI
FELLER
HAGGKVIST
CACCETTA
CAPOBIANCO

Thank You !