# Consistency of MLE and Variational Estimators in Stochastic Block Model

Alain Celisse

[1]UMR 8524 CNRS - Université Lille 1

[2]Modal INRIA team-project

[3]SSB Group, Paris

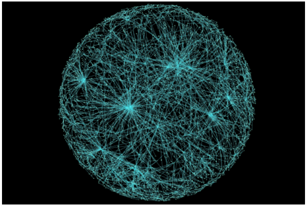
joint work with Jean-Jacques Daudin and Laurent Pierre

Workshop "Statistics for Complex Networks"

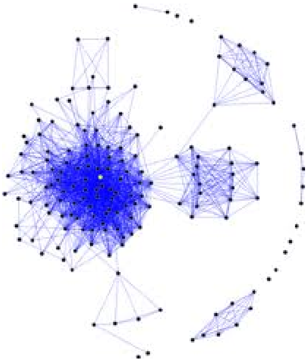
Eindhoven, January 31st, 2013

# Examples of networks

- Networks arise in numerous areas: social science, biology, internet,. . .
- Very different structures from one another.



mitochondrial metabolic network



social network

**Goal:** Get some valuable information from such networks

# Community detection

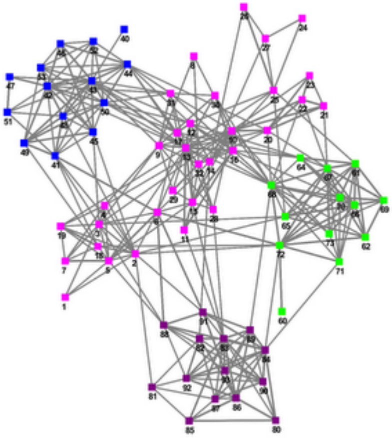
Infer relevant features:

→ to characterize the graph.

- Number of triangles, stars,. . .
- Degree of vertices,
- Connectivity of vertices.

**Goal:**

Detect subsets of vertices with the same within and between groups connectivity.



social network

# $\mathbb{I}$ Stochastic Block Model (SBM)

# Stochastic Block Model (SBM)

## Stochastic Block Model (SBM)

- $C_1, \ldots, C_Q$: $Q$ classes of vertices $\{v_1, \ldots, v_n\}$.
- $\forall 1 \leq i \leq n, \quad Z_i \in \{1, \ldots, Q\}$: label of $v_i$.

**Model:**

$$\begin{aligned} \forall 1 \leq i \leq n, \qquad & Z_i \overset{i.i.d.}{\sim} \mathcal{M}(1, \alpha_1, \alpha_2, \ldots, \alpha_Q) \ , \\ \forall 1 \leq i \neq j \leq n, \qquad & X_{i,j} \sim \mathcal{B}(\pi_{q,l}) \ , \quad \text{if } Z_i = q, Z_j = l \ , \\ & X_{i,i} = 0 \ . \end{aligned}$$

**Rk:**

$X_{i,j}$s are no longer independent.

**Goal:** Estimate

- $(\alpha_1, \ldots, \alpha_Q) \in [0,1]^Q$, with $\sum_q \alpha_q = 1$,
- $\{\pi_{q,l} \in [0,1] \mid 1 \leq q, l \leq Q\}$.

# Conditional likelihood (cSBM)

- $X_{[n]} = \{X_{i,j}\}_{1 \leq i,j \leq n}$: adjacency matrix.
- $Z_{[n]} = (Z_1, \ldots, Z_n)' \in \{1, \ldots, Q\}^n$: label vector.

Log-likelihood given $Z_{[n]} = z_{[n]}$

**Idea:**

Given $Z_{[n]} = z_{[n]}$ (labels), the $X_{i,j} \sim \mathcal{B}(\pi_{z_i, z_j})$ independent.

# Conditional likelihood (cSBM)

- $X_{[n]} = \{X_{i,j}\}_{1 \leq i,j \leq n}$: adjacency matrix.
- $Z_{[n]} = (Z_1, \ldots, Z_n)' \in \{1, \ldots, Q\}^n$: label vector.

Log-likelihood given $Z_{[n]} = z_{[n]}$

**Idea:**

Given $Z_{[n]} = z_{[n]}$ (labels), the $X_{i,j} \sim \mathcal{B}(\pi_{z_i,z_j})$ independent.

$$\mathcal{L}_1(X_{[n]}; z_{[n]}, \pi) = \log \left( \prod_{i \neq j} \pi_{z_i,z_j}^{X_{i,j}} (1 - \pi_{z_i,z_j})^{1 - X_{i,j}} \right)$$

$$= \sum_{i \neq j} X_{i,j} \log \pi_{z_i,z_j} + \sum_{i \neq j} (1 - X_{i,j}) \log \left(1 - \pi_{z_i,z_j}\right) \quad .$$

**Rk:**

Sums of independent r.v. $\longrightarrow$ concentration inequalities.

# Permutation-invariant matrices

**Example:** Let

$$\pi = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 3 & 4 \end{pmatrix} = \pi^\sigma , \quad \text{where } \sigma : \begin{matrix} 1 \mapsto 2 \\ 2 \mapsto 1 \\ 3 \mapsto 3 \end{matrix} ,$$

and $\pi^\sigma$ denote the matrix defined by $\pi^\sigma_{q,l} = \pi_{\sigma(q),\sigma(l)}$ .
Then,

$$\left\{ \pi^\sigma_{z_i,z_j} \right\}_{i,j} = \left\{ \pi_{z_i,z_j} \right\}_{i,j} = \left\{ \pi_{\sigma(z_i),\sigma(z_j)} \right\}_{i,j} .$$

**Conclusion:**

$$\longrightarrow z_{[n]} \text{ and } z^\sigma_{[n]} = \sigma\left(z_{[n]}\right) \text{ cannot be distinguished.}$$

# Permutation-invariant matrices

Permutation invariance

If there exists a permutation $\sigma(\neq Id) : \{1, \ldots, Q\} \to \{1, \ldots, Q\}$ such that

$$\pi^\sigma = \pi \quad ,$$

$\pi$ is a permutation-invariant matrix.

Equivalence classes

$$\left[\, z_{[n]} \,\right]_\pi = \left\{ z'_{[n]} \mid \exists \sigma, \quad \pi^\sigma = \pi, \quad z'_{[n]} = z^\sigma_{[n]} \right\} \quad .$$

# Concentration result in cSBM (given $Z_{[n]} = z^*_{[n]}$)

Under some assumptions detailed later, we have

**Theorem (C., Daudin, Pierre (12))**

*For every $t > 0$,*

$$P\left[\sum_{z_{[n]} \neq z^*_{[n]}} \frac{\mathbb{P}\left(Z_{[n]} = z_{[n]} \mid X_{[n]}\right)}{\mathbb{P}\left(Z_{[n]} = z^*_{[n]} \mid X_{[n]}\right)} > t \mid Z_{[n]} = z^*_{[n]}\right] = \mathcal{O}\left(ne^{-\kappa n}\right) \quad ,$$

*where $\kappa = \kappa(\pi^*) > 0$: constant and $\mathcal{O}\left(ne^{-\kappa n}\right)$ uniform w.r.t $z^*_{[n]}$.*

**Rk:**

$\rightarrow$ the same result holds without conditioning.

# Main assumptions

Assumption (A1)

$$\forall q, q', \quad \exists l \in \{1, \ldots, Q\}, \quad \pi_{q,l} \neq \pi_{q',l}, \quad \text{or} \quad \pi_{l,q} \neq \pi_{l,q'} \; .$$

**Rk:** $\rightarrow$ related to identifiability.

Assumption (A2)

$$\exists \zeta > 0, \quad \pi_{q,l} \in ]0,1[ \quad \Rightarrow \quad \pi_{q,l} \in [\zeta, 1-\zeta] \; .$$

**Rk:** $\rightarrow$ $\pi_{q,l}$s are either equal to 0 or 1, or away from 0 and 1.

Assumption (A3)

$$\exists 0 < \gamma < 1/Q, \quad \alpha_q = \mathbb{P}\left[\, Z_i = q \,\right] \in [\gamma, 1-\gamma] \; .$$

**Rk:** $\rightarrow$ No empty class.

# Generic identifiability of SBM parameters

Allman, Matias, and Rhodes (2011) proved for $Q > 2$:

$$n \text{ (even)} \geq \left( Q - 1 + \frac{(Q+2)^2}{2} \right)^2$$

implies identifiability (up to label switching) except on a set with null Lebesgue measure: *Generic Identifiability*.

**Theorem (C., Daudin, Pierre (2012))**

*Let us assume*

1. $Q \leq n/2$,
2. $\forall q,\ \alpha_q > 0$,
3. *coordinates of* $\pi \cdot \alpha$ *(or* $\pi' \cdot \alpha$*) are distinct.*

*Then, SBM parameters* $(\alpha, \pi)$ *are identifiable (up to l.s.).*

**Rk:** $\rightarrow$ (coordinates of $\pi \cdot \alpha$ (or $\pi' \cdot \alpha$) are distinct) implies (A1).

# $\mathbb{III}$ Estimation strategy

# Estimation and Computational aspects

## SBM log-likelihood

$$\mathcal{L}_2(X_{[n]};\alpha,\pi) = \log\left(\sum_{z_{[n]}} e^{\mathcal{L}_1(X_{[n]};z_{[n]},\pi)} P_{Z_{[n]}}(z_{[n]};\alpha)\right) \quad .$$

**Computational cost:**

- The sum over $z_{[n]}$ cannot be cast as a product.
- $\mathcal{L}_2(X_{[n]};\alpha,\pi)$ cannot be computed ($Q^n$ terms involved).

**Solutions:**

- MCMC $\rightarrow$ high comput. cost (Andrieu, Atchadé (2007)),
- Variational approximation (Jordan et al. (1999)).
- ...

# Variational approximation

**Idea:**

Maximize w.r.t. $(\alpha, \pi)$ a (tight) lower bound of $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$.

Variational log-likelihood:

$$\mathcal{J}(X_{[n]}; A, \alpha, \pi) = \mathcal{L}_2(X_{[n]}; \alpha, \pi) - K\left(A; P\left(Z_{[n]} = \cdot \mid X_{[n]}\right)\right) \quad ,$$

where

- $K(P, Q) \geq 0$ : Kullback-Leibler divergence between $P$ and $Q$,
- $A$: approximation to $P\left(Z_{[n]} = \cdot \mid X_{[n]}\right)$.

**Rk:**

$$\begin{aligned} &\sup_A \left\{ \mathcal{J}(X_{[n]}; A, \alpha, \pi) \right\} \\ &= \mathcal{L}_2(X_{[n]}; \alpha, \pi) - \inf_A \left\{ K\left(A; P\left[\, Z_{[n]} = \cdot \mid X_{[n]} \,\right]\right) \right\} \quad . \end{aligned}$$

# Variational approximation (Con't.)

Approximating $P\left[\, Z_{[n]} = \cdot \mid X_{[n]} \,\right]$
Choose a subset $\mathcal{R}$ of distributions such that

$$\mathcal{R} = \left\{ \prod_{i=1}^{n} \mathcal{M}(1, \tau_{i,1}, \ldots, \tau_{i,Q}) \mid \tau_{i,q} \in [0,1],\ \sum_{q=1}^{Q} \tau_{i,q} = 1 \right\} \ .$$

# Variational approximation (Con't.)

Approximating $P\left[\, Z_{[n]} = \cdot \mid X_{[n]} \,\right]$

Choose a subset $\mathcal{R}$ of distributions such that

$$\mathcal{R} = \left\{ \prod_{i=1}^{n} \mathcal{M}(1, \tau_{i,1}, \ldots, \tau_{i,Q}) \mid \tau_{i,q} \in [0,1],\ \sum_{q=1}^{Q} \tau_{i,q} = 1 \right\} \quad .$$

Final variational objective function

$$\mathcal{J}(X_{[n]}; \widehat{\tau}_{[n]}, \alpha, \pi) := \mathcal{L}_2(X_{[n]}; \alpha, \pi) - \inf_{A \in \mathcal{R}} K\left(A; P\left[\, Z_{[n]} = \cdot \mid X_{[n]} \,\right]\right) \quad .$$

**Interest:**

For every $(\alpha, \pi)$,

unlike $\mathcal{L}_2(X_{[n]}; \alpha, \pi)$, $\mathcal{J}(X_{[n]}; \widehat{\tau}_{[n]}, \alpha, \pi)$ can be computed.

# How to prove consistency of VE?

Maximum likelihood estimator (MLE)

$$(\hat{\alpha}, \hat{\pi}) := \mathrm{Argmin}_{(\alpha,\pi)} \mathcal{L}_2(X_{[n]}; \alpha, \pi) \ .$$

Variational estimator (VE)

$$(\tilde{\alpha}, \tilde{\pi}) := \mathrm{Argmin}_{(\alpha,\pi)} \mathcal{J}(X_{[n]}; \widehat{\tau}_{[n]}, \alpha, \pi) \ .$$

**Main goal:**

Settle consistency of Variational Estimator (VE).

# $\mathbb{III}$ Main results

# Uniform convergence

**Theorem (C., Daudin, Pierre (12))**

*Under assumptions* $(A1)-(A3)$,

$$\sup_{(\alpha,\pi)} \frac{1}{n(n-1)} \left| \mathcal{J}(X_{[n]}; \widehat{\tau}_{[n]}, \alpha, \pi) - \mathcal{L}_2(X_{[n]}; \alpha, \pi) \right| \xrightarrow[n\to+\infty]{a.s.} 0 \ .$$

**Idea:**

MLE and VE share the same asymptotic properties.

**Two-step Strategy:**

1. Settle consistency of MLE.
2. Deduce consistency of VE.

# Different scaling between $\alpha$ and $\pi$

## Log-likelihood of SBM

$$\mathcal{L}_2(X_{[n]};\alpha,\pi) = \log\left(\sum_{z_{[n]}} e^{\mathcal{L}_1(X_{[n]};z_{[n]},\pi)} P_{Z_{[n]}}(z_{[n]};\alpha)\right) ,$$

with

- $P_{Z_{[n]}}(z_{[n]};\alpha) = \prod_{q=1}^{Q}\left(\alpha_q^{\sum_{i=1}^n 1_{(z_i=q)}}\right)$ ,
- $\mathcal{L}_1(X_{[n]};z_{[n]},\pi) = \sum_{q,l}\left[S_{q,l}(z_{[n]})\log\pi_{q,l} + \left(N_{q,l}(z_{[n]}) - S_{q,l}(z_{[n]})\right)\log\left(1-\pi_{q,l}\right)\right]$ ,

  where

$$N_{q,l}(z_{[n]}) = \sum_{i\neq j} \mathbb{1}_{(z_i=q,z_j=l)},$$
$$S_{q,l}(z_{[n]}) = \sum_{i\neq j} X_{i,j}\mathbb{1}_{(z_i=q,z_j=l)}.$$

**Idea:**

$\longrightarrow$ Two-step estimation strategy: Estimate first $\pi$, then $\alpha$.

# MLE consistency: General consistency theorem

## Theorem (C., Daudin, Pierre (12))

*Let* $M_n : \ \Theta \times \Psi \to \mathbb{R}$*: random function and* $\mathbb{M} : \ \Theta \to \mathbb{R}$ *deterministic such that for every* $\epsilon > 0$*,*

$$\sup_{d(\theta,\theta_0)\geq\epsilon} \mathbb{M}\left(\theta\right) < \mathbb{M}\left(\theta_0\right) \quad ,$$

$$\sup_{(\theta,\psi)} \left|M_n\left(\theta,\psi\right) - \mathbb{M}\left(\theta\right)\right| \xrightarrow[n\to+\infty]{P} 0 \quad .$$

*If* $(\widehat{\theta},\widehat{\psi}) = \mathrm{Argmax}_{\theta,\psi} M_n\left(\theta,\psi\right)$*, then*

$$d\left(\widehat{\theta},\theta_0\right) \xrightarrow[n\to+\infty]{P} 0 \quad .$$

**Remark:**

$$\theta = \pi,\ \psi = \alpha, \quad M_n(\pi,\alpha) = [n(n-1)]^{-1}\mathcal{L}_2(X_{[n]};\alpha,\pi).$$

# MLE consistency: Main strategy

First step: $P(\cdot \mid Z_{[n]} = z^*_{[n]})$ as reference probability.
Concentration inequalities with

$$\sum_{r,t} \sum_{i \neq j, z_i = q, z_j = l} X_{i,j} \mathbb{1}_{(z_i^* = r, z_j^* = t)} \quad \text{(ind. r.v.)} \quad .$$

For every $u > 0$,

$$\left| \sum_{r,t} \sum_{i \neq j, z_i = q, z_j = l} (X_{i,j} - \pi^*_{r,t}) \mathbb{1}_{(z_i^* = r, z_j^* = t)} \right| > u$$

with proba. $\leq \exp\left[ -\Delta \cdot N_{q,l}(z_{[n]}) \cdot u^2 \right]$ ($\Delta$: constant).

# MLE consistency: Main strategy

First step: $P(\cdot \mid Z_{[n]} = z^*_{[n]})$ as reference probability.
Concentration inequalities with

$$\sum_{r,t} \sum_{i \neq j, z_i = q, z_j = l} X_{i,j} \mathbb{1}_{(z_i^* = r, z_j^* = t)} \quad \text{(ind. r.v.)} \ .$$

For every $u > 0$,

$$\left| \sum_{r,t} \sum_{i \neq j, z_i = q, z_j = l} (X_{i,j} - \pi^*_{r,t}) \mathbb{1}_{(z_i^* = r, z_j^* = t)} \right| > u$$

with proba. $\leq \exp\left[ -\Delta \cdot N_{q,l}(z_{[n]}) \cdot u^2 \right]$ ($\Delta$: constant).

Second step: Integrating over $Z_{[n]}$

Concentration of $\sum_{i=1}^n (\mathbb{1}_{(Z_i = q)} - \alpha_q)$ (ind. Bernoulli r.v.).

# MLE consistency

**Theorem (C., Daudin, Pierre (12))**

$$(\widehat{\alpha}, \widehat{\pi}) := \mathrm{Argmax}_{(\alpha,\pi)} \mathcal{L}_2(X_{[n]}; \alpha, \pi) \ .$$

*Then for any distance $d(\cdot,\cdot)$ on the set of parameters $\pi$,*

$$d\left(\widehat{\pi}, \pi^*\right) \xrightarrow[n \to +\infty]{\mathbb{P}} 0 \ .$$

**Theorem (C., Daudin, Pierre (12))**

*If* $\|\widehat{\pi} - \pi^*\|_\infty = o_{\mathbb{P}}\left(\sqrt{\log n}/n\right)$, *then*

$$d(\widehat{\alpha}, \alpha^*) \xrightarrow[n \to +\infty]{\mathbb{P}} 0 \ .$$

$\longrightarrow$ Similar results for VE.

# Take-home message and prospects

## Recap.

- Main tool: concentration inequalities.
- Two-step strategy: ($i$) cSBM, and ($ii$) SBM.
- First theoret. guarantees on consistency of VE (MLE) in SBM.

## Prospects

- Rates of convergence of MLE and VE estimators.
- Distinguish between regimes (let $\zeta$ and $\gamma \to 0$):
  $\mathbb{P}[X_{i,j} = 1 \mid Z_i, Z_j] \to 0$ as $n \to +\infty$
  (Bickel, Choi, Chang, Zhang (12)).

# Take-home message and prospects

Recap.

- Main tool: concentration inequalities.
- Two-step strategy: (*i*) cSBM, and (*ii*) SBM.
- First theoret. guarantees on consistency of VE (MLE) in SBM.

Prospects

- Rates of convergence of MLE and VE estimators.
- Distinguish between regimes (let $\zeta$ and $\gamma \to 0$): $\mathbb{P}\,[\,X_{i,j} = 1 \mid Z_i, Z_j\,] \to 0$ as $n \to +\infty$ (Bickel, Choi, Chang, Zhang (12)).

Thank you!

# Bibliography

SBM

- Holland, P., Laskey, K., and Leinhardt, S. (1983). Stochastic blockmodels: Some first steps. Social Networks, 5, 109–137
- Nowicki, K. and Snijders, T. (2001). Estimation and prediction for stochas- tic block-structures. J. Am. Stat. Assoc., 96, 1077–1087
- Daudin, J.-J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. Stat Comput , 18, 173–183.
- Ambroise, C. and Matias, C. (2012). New consistent and asymptotically normal estimators for random graph mixture models. JRSS. B, 74(1), 3–35

cSBM

- Bickel, P. and Chen, A. (2009). A nonparametric view of network models and Newman-Girvan and other modularities. PNAS, pages 1–6
- Choi, D. S., Wolfe, P. J., and Airoldi, E. M. (2012). Stochas- tic blockmodels with growing number of classes. Biometrika
- Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. Ann. Statist., 39(4), 1878–1915