Confidence sets in a sparse stochastic block model with two communities of unknown sizes

arXiv:2108.07078 [math.ST]

Bas Kleijn
KdV Institute for Mathematics

Universiteit van Amsterdam
ABSTRACT

In a sparse stochastic block model with two communities of unequal sizes we derive two posterior concentration inequalities, for (1) posterior (almost-)exact recovery of the community structure; (2) a construction of confidence sets for the community assignment from credible sets with finite graph sizes, enabling exact frequentist uncertain quantification with Bayesian credible sets at non-asymptotic graph sizes. It is argued that a form of early stopping applies to MCMC sampling of the posterior to enable the computation of confidence sets at larger graph sizes.

[Based on joint work with J. van Waaij]


B. Kleijn, J. van Waaij, arxiv:1810.09533, 2108.07078
Part I

Sparse stochastic block models
Erdös-Rényi random graphs

Fix $n \geq 1$, denote $G_n = (V_n, E_n)$ complete graph with $n$ vertices and percolate edges,

For every $e \in E_n$ independently, include $e$ in $E'_n \subset E_n$ wp. $p_n$.

Result random graph $G(n, p_n) = (V_n, E'_n)$ (Erdös, Rényi (1959–1961)).
Sparsity phases of the Erdös-Rényi random graph

Fragmented

\[ p_n < \frac{1}{n} \]
Many fragments
clusters \( \leq O(\log(n)) \)
\[ E(N_i) = O(1) \]

Kesten-Stigum

\[ \frac{1}{n} < p_n = a_n/n < \frac{\log(n)}{n} \]
Giant component
cluster \( \sim O(n) \)
\[ E(N_i) = O(a_n) \]

Chernoff-Hellinger

\[ p_n > \frac{\log(n)}{n} \]
Connected
cluster = \( n \)
\[ E(N_i) = O(\log(n)) \]
Two-community stochastic block model

Consider $G_n = (V_n, E_n)$ with community assignment $\theta_n \in \Theta_n = \{0, 1\}^n$. Split $V_n = Z_0(\theta_n) \cup Z_1(\theta_n)$. For every $e \in E_n$ independently,

\[
\begin{cases} 
    p_n, & \text{if } e \text{ lies within } Z_0 \text{ or } Z_1, \\
    q_n, & \text{if } e \text{ lies between } Z_0 \text{ and } Z_1. 
\end{cases}
\]

Three-community SBM graph $X^n = (V_n, E'_n) \in \mathcal{X}_n$, $X^n \sim P_{\theta_n}$
Community detection

Example SBM with $n = 12$, $0 < q_n \ll p_n < 1$, $\theta_n = 000000111111$

Observation
Data $X^n \sim P_{\theta_n}$

Unobserved
Communities of $\theta_n$
$Z_0(\theta_n), Z_1(\theta_n)$

Detection
Estimate with
$\hat{Z}_0(X^n), \hat{Z}_1(X^n)$
Asymptotic community detection

**Definition 8.1** Given community assignments \( \theta_n \) for all \( n \geq 1 \), an estimator sequence \( \hat{\theta}_n : \mathcal{X}_n \rightarrow \Theta_n \) is said to recover \( \theta_n \) exactly, if,

\[
P_{\theta_n} \left( \hat{\theta}_n(X^n) = \theta_n \right) \rightarrow 1,
\]
as \( n \rightarrow \infty \).

Let \( k : \Theta_n \times \Theta_n \rightarrow \{0, 1, \ldots, n\} \) denote the Hamming distance.

**Definition 8.2** Given community assignments \( \theta_n \) for all \( n \geq 1 \) and some sequence of error rates \( (k_n) \) of order \( k_n = O(n) \), an estimator sequence \( \hat{\theta}_n : \mathcal{X}_n \rightarrow \Theta_n \) is said to recover \( \theta_n \) almost-exactly with error rate \( k_n \), if,

\[
P_{\theta_n} \left( k(\hat{\theta}_n(X^n), \theta_n) \leq k_n \right) \rightarrow 1,
\]
as \( n \rightarrow \infty \).
Part II

Posterior concentration
Posterior concentration (I)

Let,

\[ \rho(p, q) = p^{1/2}q^{1/2} + (1 - p)^{1/2}(1 - q)^{1/2}, \]

denote the Hellinger-affinity between two Bernoulli-distributions with parameters \( p, q \in (0, 1) \).

**Theorem 10.1** For fixed \( n \geq 1 \), suppose \( X^n \sim P_{\theta_n} \) with \( \theta_n \in \Theta_n \) and choose the uniform prior on \( \Theta_n \). Then,

\[ E_{\theta_n}\cap(\{\theta_n\}|X^n) \geq 1 - \frac{n}{2}\rho(p_n, q_n)^{n/2} e^{n\rho(p_n, q_n)^{n/2}}, \]

implying that if,

\[ n\rho(p_n, q_n)^{n/2} \to 0, \quad (1) \]

then the posterior recovers the true community assignment exactly.
Exact recovery in the Chernoff-Hellinger phase

Sparsity

\[ p_n = a_n \frac{\log(n)}{n}, \quad q_n = b_n \frac{\log(n)}{n}. \]

**Corollary 11.1** Assume the conditions of theorem 10.1. If the sequences \(a_n, b_n\) in the Chernoff-Hellinger phase satisfy,

\[
\left( (\sqrt{a_n} - \sqrt{b_n})^2 - \frac{a_nb_n \log(n)}{2n} - 4 \right) \log(n) \to \infty,
\]

then the posterior recovers the community assignments exactly.

For \(a_n, b_n\) of order \(O(1)\), a simple sufficient conditions for exact recovery is,

\[
\left( (\sqrt{a_n} - \sqrt{b_n})^2 - 4 \right) \log n \to \infty,
\]
Posterior concentration (II)

Define the (Hamming-)metric balls,

\[ B_n(\theta_n, k_n) = \{ \eta_n \in \Theta_n : k(\eta_n, \theta_n) \leq k_n \} \],

(4)

**Theorem 12.1** For fixed \( n \geq 1 \), suppose \( X^n \sim P_{\theta_n} \) with \( \theta_n \in \Theta_n \) and choose the uniform prior on \( \Theta_n \). For some \( \lambda_n \) with \( 0 < \lambda_n < 1/2 \), let \( k_n \) be an integer such that \( k_n \geq \lambda_n n \). Then,

\[
E_{\theta_n} \Pi( B_n(\theta_n, k_n) \mid X^n ) \geq 1 - \frac{1}{2} \left( \frac{e^{-1}}{\lambda_n \rho(p_n, q_n)^{n/2}} \right)^{\lambda_n n} \left( 1 - \frac{e^{-1}}{\lambda_n \rho(p_n, q_n)^{n/2}} \right)^{-1}.
\]
Recovery in the Kesten-Stigum phase (I)

Sparsity \( p_n = \frac{c_n}{n}, \quad q_n = \frac{d_n}{n} \).

**Proposition 13.1** If the sequences \( c_n, d_n \) and the fractions \( \lambda_n \) satisfy,
\[
\lambda_n n \left( \log(\lambda_n) + \frac{1}{4} \left( \sqrt{c_n} - \sqrt{d_n} \right)^2 - 1 \right) \to \infty,
\]
(5) then posteriors recover the community assignment almost-exactly with any error rate \( k_n \geq \lambda_n n \).

**Corollary 13.2** Recovery c.f. (Decelle et al. (2011))

Let \( 0 < \lambda < 1/2 \) be given. If, for some constant \( C > 1 \) and large enough \( n \),
\[
(\sqrt{c_n} - \sqrt{d_n})^2 > 4C(1 - \log(\lambda)),
\]
(6) then the posterior recovers the community assignment almost exactly with error rate \( k_n = \lambda n \).
Recovery in the Kesten-Stigum phase (II)

Corollary 14.1 Weak consistency (Mossel, Neeman, Sly (2016))
If the sequences $c_n$ and $d_n$ satisfy,

$$\frac{(c_n - d_n)^2}{2(c_n + d_n)} \to \infty,$$

(7)

the posterior recovers the true community assignment almost exactly with any error rate $k_n \geq \lambda_n n$ for some vanishing fraction $\lambda_n \to 0$.

Corollary 14.2 Let $0 < \lambda_n < 1/2$ be given, such that $\lambda_n \to 0$, $\lambda_n n \to \infty$. If, for some constant $C > 1$ and large enough $n$,

$$(\sqrt{c_n} - \sqrt{d_n})^2 + 4C \log(\lambda_n) \to \infty,$$

(8)

then the posterior recovers the community assignments almost exactly with error rate $k_n = \lambda_n n$. 

14
Part III

Uncertainty quantification
Bayesian and frequentist uncertainty quantified

**Definition 16.1**  Given \( n \geq 1 \), a prior \( \Pi_n \) and data \( X^n \), a **credible set** of credible level \( 1 - \gamma \) is any \( D(X^n) \subset \Theta_n \) such that:

\[
\Pi(D(X^n)|X^n) \geq 1 - \gamma,
\]

\( P_{\Pi_n} \)-almost-surely. In case \( \gamma = 0 \), \( D(X^n) \) is the support of the posterior.

**Definition 16.2**  Given \( \theta_n \in \Theta_n \) and data \( X^n \sim P_{\theta_n} \), a **confidence set** \( C(X^n) \subset \Theta_n \) of confidence level \( 1 - \alpha \) is defined by any \( x^n \mapsto C(x^n) \subset \Theta_n \) such that,

\[
P_{\theta_n}(\theta_n \in C(X^n)) \geq 1 - \alpha.
\]
Lemma 17.1 Fix $n \geq 1$, let $\theta_n \in \Theta_n$, $X^n \sim P_{\theta_n}$ be given. For any $B \subset \Theta_n$, $0 < \beta < 1$,

$$E_{\theta_n} \Pi(B|X^n) \geq 1 - \beta \implies P_{\theta_n}(B \cap D(X^n) \neq \emptyset) \geq 1 - \frac{\beta}{1 - \gamma}.$$ 

for any credible set $D(X^n) \subset \Theta_n$ of credible level $1 - \gamma$.

![Diagram showing the enlargement of set $D$ by sets $B(\theta)$ to form set $C$.](image)
Credible sets are confidence sets (I)

**Proposition 18.1** For fixed $n \geq 1$, suppose $X^n \sim P_{\theta_n}$ with $\theta_n \in \Theta_n$. Every credible set $D(X^n)$ of credible level $1 - \gamma$ is a confidence set of confidence level,

$$P_{\theta_n}(\theta_n \in D(X^n)) \geq 1 - \frac{n}{2(1 - \gamma)}\rho(p_n, q_n)^{n/2} e^{n\rho(p_n, q_n)^{n/2}}. \quad (9)$$

**Method 18.2** For graph size $n$, realised graph $X^n = x^n$, known $p, q$ and realised posterior $\Pi(\cdot | X^n = x^n)$, choose a desired confidence level $0 < 1 - \alpha < 1$, we choose credible level,

$$1 - \gamma = \min\{1, (n/2\alpha)\rho(p, q)^{n/2} e^{n\rho(p, q)^{n/2}}\}. \quad (10)$$
Example 19.1  Take $p = 0.9$, $q = 0.1$ and confidence level $1 - \alpha = 0.95$. $\rho(p, q) = 0.6$ and $(n/2)\rho(p, q)^{n/2} \approx 0.0211$. As $n$ varies,

any (unenlarged) credible set of credible level $1 - \gamma$ is a confidence set of confidence level $0.95$
Enlarged credible sets are confidence sets (I)

The $k$-enlargement $C(X^n)$ of $D(X^n)$ is the union of all Hamming balls of radius $k \geq 1$ that are centred on points in $D(X^n)$,

$$C(X^n) = \left\{ \theta_n \in \Theta_n : \exists \eta_n \in D_n(X^n), k(\theta_n, \eta_n) \leq k \right\},$$

**Proposition 20.1** For fixed $n \geq 1$, suppose $X^n \sim P_{\theta_n}$ with $\theta_n \in \Theta_n$. Define $k = \lceil \lambda n \rceil$. Then the $k$-enlargement $C(X^n)$ of any credible set $D(X^n)$ of level $1 - \gamma$ is a confidence set of confidence level,

$$P_{\theta_n}(\theta_n \in C(X^n)) \geq 1 - \frac{1}{2(1 - \gamma)} \left( \frac{e \rho(p_n, q_n)^{n/2}}{\lambda n} \right)^{\lambda n} \left( 1 - \frac{e \rho(p_n, q_n)^{n/2}}{\lambda n} \right)^{-1}.$$
Enlarged credible sets are confidence sets (II)

Example 21.1  Again $p = 0.9$, $q = 0.1$ and confidence level $1 - \alpha = 0.95$. For $\lambda = 0.05$ and varying graph size $n$,

any $0.05n$-enlarged credible set of credible level $1 - \gamma$ is also a confidence set of confidence level 0.95.

Required credible level for confidence level $1 - \alpha = 0.95$ ($\lambda = 0.05$)

\[ n \]
Enlarged credible sets are confidence sets (III)

Example 22.1  Again $p = 0.9$, $q = 0.1$ and confidence level $1 - \alpha = 0.95$. For $\lambda = 0.1$ and varying graph size $n,$

any $0.1n$-enlarged credible set of credible level $1 - \gamma$ is also a confidence set of confidence level $0.95$

![Diagram](image_url)
Enlarged credible sets are confidence sets (IV)

Example 23.1  Again \( p = 0.9, \ q = 0.1 \) and confidence level \( 1 - \alpha = 0.95 \). For \( \lambda = 0.25 \) and varying graph size \( n \),

any \( 0.25n \)-enlarged credible set of credible level \( 1 - \gamma \) is also a confidence set of confidence level 0.95.

![Graph showing the relationship between \( n \) and \( 1 - \gamma \).

Required credible level for confidence level \( 1 - \alpha = 0.95 \) \( (\lambda = 0.25) \).
Part IV

Asymptotic uncertainty quantification
Asymptotic credible and confidence sets

**Definition 25.1** Let \((\Theta, \mathcal{G})\) with priors \(\Pi_n\) and a collection \(\mathcal{D}\) of measurable subsets of \(\Theta\) be given. **Credible sets** \((D_n)\) of credible levels \(1 - o(a_n)\) are maps \(D_n : \mathcal{X}_n \rightarrow \mathcal{D}\) such that,

\[
P(\Theta \setminus D_n(X^n) | X^n) = o(a_n),
\]

\(P_{\Pi_n}\)-almost-surely.

**Definition 25.2** Maps \(x \mapsto C_n(x) \subset \Theta\) are asymptotically consistent confidence sets (of levels \(1 - o(a_n)\)), if,

\[
P_{\theta,n}(\theta \notin C_n(X^n)) \rightarrow 0, \quad (= o(a_n))
\]

for all \(\theta \in \Theta\). \(C_n\) is asymptotically informative, if for all \(\theta' \neq \theta\),

\[
P_{\theta',n}(\theta \in C_n(X^n)) \rightarrow 0
\]
Credible sets with converging posteriors

**Theorem 26.1** Suppose that $0 < \epsilon \leq 1$, $P_{\theta_0,n} \ll P_n^{\Pi_n}$ and

\[
\Pi \left( d_n(\theta_n, \theta_{0,n}) \leq r_n \mid X^n \right) \xrightarrow{P_{\theta_0,n}} 1
\]

Let $\hat{D}_n(X^n) = B_n(\hat{\theta}_n, \hat{r}_n)$ be level-$1 - \epsilon$ credible balls of minimal radii.

Then with high $P_{\theta_0,n}$-probability $\hat{r}_n \leq r_n$ and the sets,

\[
C_n(X^n) = B_n(\hat{\theta}_n, \hat{r}_n + r_n) \subset B_n(\hat{\theta}_n, 2r_n)
\]

have asymptotic coverage,

\[
P_{\theta_0,n} \left( \theta_{0,n} \in C_n(X^n) \right) \to 1,
\]
Credible sets \textit{without} converging posteriors

\textbf{Theorem 27.1} Let $0 \leq a_n \leq 1$, $a_n \downarrow 0$ and $b_n > 0$ such that $a_n = o(b_n)$ be given and let $D_n$ denote level-$(1 - a_n)$ credible sets. Furthermore, for all $\theta \in \Theta$, let $B_n$ be set functions such that,

(i) $\Pi_n(B_n(\theta_0)) \geq b_n,$

(ii) $P_{\theta_0,n} < b_n a_n^{-1} P_n | B_n(\theta_0).$

Then the credible sets $D_n$, enlarged by the sets $B_n$, are \textit{asymptotically consistent} confidence sets $C_n$, that is,

$$P_{\theta_0,n}(\theta_0 \in C_n(X^n)) \rightarrow 1.$$
Discussion

**Sharpness of the bounds** If posterior concentration bounds are not sharp, lower bounds for credible levels become unnecessary high and enlargement radii become unnecessarily large.

**Early stopping** Since only community assignments with high posterior probabilities are needed in credible sets of low credible level, small MCMC samples may not hamper the construction of confidence sets: some form of early stopping of the MCMC sequence may be justified.

**Generalization and cross validation** All of this generalizes and can be verified by simulation and cross validation.

Thank you for your attention

arXiv:2108.07078 [math.ST]
Extra Remote contiguity
Remote contiguity

**Definition 30.1** Given $(P_n), (Q_n)$, $Q_n$ is contiguous w.r.t. $P_n$ ($Q_n \triangleleft P_n$), if for any msb $\psi_n : \mathcal{X}^n \rightarrow [0, 1]$

$$P_n \psi_n = o(1) \Rightarrow Q_n \psi_n = o(1)$$

**Definition 30.2** Given $(P_n), (Q_n)$ and a $a_n \downarrow 0$, $Q_n$ is $a_n$-remotely contiguous w.r.t. $P_n$ ($Q_n \triangleleft a_n^{-1} P_n$), if for any msb $\psi_n : \mathcal{X}^n \rightarrow [0, 1]$

$$P_n \psi_n = o(a_n) \Rightarrow Q_n \psi_n = o(1)$$

**Remark 30.3** Contiguity is stronger than remote contiguity.

Note that $Q_n \triangleleft P_n$ iff $Q_n \triangleleft a_n^{-1} P_n$ for all $a_n \downarrow 0$. 
Le Cam’s first lemma

**Lemma 31.1** Given \((P_n), (Q_n)\) like above, \(Q_n \prec P_n\) iff:

(i) If \(T_n \xrightarrow{P_n} 0\), then \(T_n \xrightarrow{Q_n} 0\)

(ii) Given \(\epsilon > 0\), there is a \(b > 0\) such that \(Q_n(\frac{dQ_n}{dP_n} > b) < \epsilon\)

(iii) Given \(\epsilon > 0\), there is a \(c > 0\) such that \(\|Q_n - Q_n \wedge c P_n\| < \epsilon\)

(iv) If \(dP_n/dQ_n \xrightarrow{Q_n-w.} f\) along a subsequence, then \(P(f > 0) = 1\)

(v) If \(dQ_n/dP_n \xrightarrow{P_n-w.} g\) along a subsequence, then \(Eg = 1\)
Criteria for remote contiguity

Lemma 32.1 Given \((P_n), (Q_n)\), \(a_n \downarrow 0\), \(Q_n \prec a_n^{-1} P_n\) if any of the following holds:

(i) For any bnd msb \(T_n : \mathcal{X}^n \rightarrow \mathbb{R}\), \(a_n^{-1} T_n \xrightarrow{P_n} 0\), implies \(T_n \xrightarrow{Q_n} 0\)

(ii) Given \(\epsilon > 0\), there is a \(\delta > 0\) s.t. \(Q_n(\frac{dP_n}{dQ_n} < \delta a_n) < \epsilon\) f.l.e.n.

(iii) There is a \(b > 0\) s.t. \(\liminf_{n \to \infty} b a_n^{-1} P_n(\frac{dQ_n}{dP_n} > b a_n^{-1}) = 1\)

(iv) Given \(\epsilon > 0\), there is a \(c > 0\) such that \(\|Q_n - Q_n \wedge c a_n^{-1} P_n\| < \epsilon\)

(v) Under \(Q_n\), every subsequence of \((a_n(\frac{dP_n}{dQ_n})^{-1})\) has a weakly convergent subsequence