Asymptotic uncertainty quantification for communities in sparse planted bi-section models

Kleijn, B. J.K.; van Waaij, J.

DOI
10.1016/j.jspi.2023.04.002

Publication date
2023

Document Version
Final published version

Published in
Journal of Statistical Planning and Inference

Journal of Statistical Planning and Inference 227 (2023) 112–128

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Asymptotic uncertainty quantification for communities in sparse planted bi-section models

B.J.K. Kleijn [a], J. van Waaij [b,*]

[a] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
[b] Department of Health Technology, Section of Bioinformatics, Technical University of Denmark, Kemitorvet, 204, 252 2800 Kongens Lyngby, Denmark

ARTICLE INFO

Article history:
Received 1 April 2022
Received in revised form 13 April 2023
Accepted 21 April 2023
Available online 29 April 2023

MSC:
05C80
62C10
62G05
62G15

Keywords:
Community detection
Sparse random graph
Posterior consistency
uncertainty quantification

ABSTRACT

Posterior distributions for community structure in sparse planted bi-section models are shown to achieve exact (resp. almost-exact) recovery, with sharp bounds for the sparsity regimes where edge probabilities decrease as $O(\log(n)/n)$ (resp. $O(1/n)$). Assuming posterior recovery, one may interpret credible sets (resp. enlarged credible sets) as asymptotically consistent confidence sets; the diameters of those credible sets are controlled by the rate of posterior concentration. If credible levels are chosen to grow to one quickly enough, corresponding credible sets can be interpreted as frequentist confidence sets without conditions on posterior concentration. In the regimes with $O(1/n)$ edge sparsity, or when within-community and between-community edge probabilities are very close, credible sets may be enlarged to achieve frequentist asymptotic coverage, also without conditions on posterior concentration.

1. Community detection and uncertainty quantification

One of the central questions in network science concerns community detection (Girvan and Newman, 2002): one observes a graph with vertices that belong to various (unobserved) communities and edges that are present or not with community-dependent probabilities. The goal is to infer the community structure based on the presence or absence of edges in the observed graph. The stochastic block model (Holland et al., 1983) is the most popular model for the observation: edges in the observed random graph occur independently, with probabilities that depend on the community membership of the vertices they connect. As such, the stochastic block model is an inhomogeneous generalization of the Erdős-Rényi model (Erdős and Rényi, 1959). These days stochastic block models are applied in all branches of science and its applications and are widely employed as canonical models for the study of clustering and community structure (Fortunato, 2010; Abbe, 2018).

Aside from its applications, the theory of the community detection problem has attracted great attention from outside network science, particularly from statistical physics, machine learning, probability theory, combinatorics and statistics. The machine learners' practically oriented perspective has led to a wide range of algorithms for community detection, in which computability is central. From the more stochastically centred perspective of probabilists and statisticians, a large variety of estimation methods for community structure has been proposed, including spectral clustering (see Krzakala et al.

* Corresponding author.
E-mail address: jvawa@dtu.dk (J. van Waaij).

https://doi.org/10.1016/j.jspi.2023.04.002

(2013) and many others), maximization of the likelihood and other modularities (Girvan and Newman, 2002; Bickel and Chen, 2009; Choi et al., 2012; Amini et al., 2013), semi-definite programming (Hajek et al., 2016; Guédon and Vershynin, 2016), and penalized ML detection of communities with minimax optimal misclassification ratio (Zhang and Zhou, 2016; Gao et al., 2017). Bayesian methods have been popular throughout, *e.g.*, the original work (Nowicki and Snijders, 2001), the work of Decelle et al. (2011b,a), with uniform priors and, more recently, Suwan et al. (2016) with an empirical prior choice. MCMC simulation of posterior distributions for community structure is discussed, for example, in McDaid et al. (2013), Geng et al. (2019) and Jiang and Tokdar (2021). This very brief summary does not do justice to the vast size and enormous variety of the literature on community detection methods, and we refer to the highly informative review of Abbe (2018) for an extensive bibliography and a more comprehensive discussion.

Community detection in very large graphs is used to assess and compare detection methods: large numbers of edges supply large amounts of information on community structure and community detection methods should therefore be more accurate in large graphs. Asymptotically, a natural requirement for any detection method is consistency: as the number of vertices goes to infinity, community estimates are required to coincide with (exact recovery) or converge to (almost-exact recovery) the true, underlying community structure with high probability. Let us denote the observed graph by X^n, where n is the number of vertices; the (random) presence (resp. absence) of an edge between vertices labelled $1 \leq i, j \leq n$, $i \neq j$ is denoted $X^n_{ij} = 1$ (resp. $X^n_{ij} = 0$). In most variations of the stochastic block model, the vertices belong exclusively to one of $K \geq 2$ communities as described by the unobserved community assignment vector θ_n with components $\theta_i \in \{0, \ldots, K-1\}$, $(1 \leq i \leq n)$. Community k, $(0 \leq k \leq K-1)$, has $n_k = \sum_i 1_{\theta_i = k}$ members and vertex i, $(1 \leq i \leq n)$, has (random) degree $\sum_j X^n_{ij}$. The edge connecting vertices i and j occurs with probability $Q_n(\theta_i, \theta_j)$ depending on the communities of those vertices. Stochastic block models vary in that they assume known or unknown: the number of communities K, the edge probabilities $Q_n : \{0, \ldots, K-1\}^2 \to [0, 1]$ and/or the sizes $n_0, \ldots, n_{K-1}$ of the communities.

Note that the expected degree of vertex i is equal to $\sum_{j \neq i} Q_n(\theta_i, \theta_j)$, implying that with n-independent edge-probabilities expected degrees are proportional to the graph size n. Many proofs of asymptotic consistency for community detection methods are based on models in which Q_n does not depend on the graph size n, meaning that they describe limits of stochastic block models *with unbounded degrees*. Real-world networks (*e.g.*, social networks or citation networks) describe vertices with expected degrees that stay bounded or grow more slowly with the size of the network. To describe large networks with bounded or slowly-growing degrees, a form of *edge sparsity* is required: edge probabilities must decrease with increasing graph size (this point is also emphasized in Yuan et al. (2022)). For example, if $Q_n = O(n^{-1})$, expected degrees are bounded, and if $Q_n = O(n^{-1}\log(n))$, expected degrees grow logarithmically with n. One may then wonder *which levels of edge sparsity make community detection only just possible*; or conversely, at which level of edge sparsity do community detection and other forms of inference on the community structure become impossible?

In Dyer and Frieze (1989), Decelle et al. (2011b,a), Abbe et al. (2016), Massoulié (2014), Mossel et al. (2016) and many other publications, feasibility of the community detection problem and sharp bounds on edge sparsity are studied in the context of the so-called *planted bi-section model*, which is a stochastic block model with $K = 2$ equally-sized communities of n vertices each and edge probabilities p_n (within communities) and q_n (between communities) that decrease with n (for a more detailed description of the model, see Section 2). The answers relate to the three sparsity phases (*i.e.*, fragmented, giant-component or connected) of the Erdős-Rényi graph (Erdős and Rényi, 1959; Bollobás et al., 2007): for example, Dyer and Frieze (1989) showed that minimization of the number of edges between estimated communities finds the true community assignment vector with high probability, if there exists a constant $A > 0$ such that, $p_n - q_n \geq An^{-1}\log n$; in Mossel et al. (2016) it is shown that community detection with errors that converge to zero in probability is possible, if and only if,

$$\frac{n(p_n - q_n)^2}{p_n + q_n} \to \infty, \tag{1}$$

(see also Decelle et al., 2011b,a). In Massoulié (2014), Abbe et al. (2016) and Mossel et al. (2015, 2016) it is shown that if we write $p_n = a_n n^{-1}\log n$ and $q_n = b_n n^{-1}\log n$, assuming that $C^{-1} \leq a_n, b_n \leq C$, estimates coinciding with the true community assignment vector with high probability are possible, if and only if,

$$(a_n + b_n - 2\sqrt{a_n b_n} - 1)\log n + \tfrac{1}{2}\log\log n \to \infty. \tag{2}$$

Conditions (1) and (2) not only lower-bound the degree of edge-sparsity, but also guarantee sufficient distinction (Janson, 2010; Banerjee, 2018) from the Erdős-Rényi graph ($p_n = q_n$), in which communities are not identifiable.

Besides community detection, other forms of inference on the parameters defining a stochastic block model are studied. For example, Bickel and Sarkar (2016) and Lei (2016) define asymptotically consistent tests for the number of communities in a stochastic block models with unbounded degrees. In Yuan et al. (2022) an asymptotically consistent likelihood ratio test is considered to distinguish between the Erdős-Rényi graph in a planted bi-section graph with bounded degrees.

The first goal of this paper is to explore the behaviour of posterior distributions for the community assignment vector in the planted bi-section model with bounded and slowly-growing degrees and to demonstrate appropriate forms of *posterior consistency* under condition (1) and (a slight variation on) condition (2) (see Section 3). The second, more important goal is *frequentist uncertainty quantification* based on an advantage that posteriors offer over other estimation methods: in

Section 4, Bayesian credible sets for community assignment are shown to be (or can be enlarged to form) *asymptotically consistent confidence sets*. In Section 5 we draw conclusions, discuss some further possibilities and relate to other work.

Section 2 introduces the planted bi-section model, the Bayesian posterior and test functions to prove its convergence. Appendix A establishes notation and basic Bayesian definitions; Appendix B introduces remote contiguity and applies it to convert credible sets to confidence sets, as in Kleijn (2021).

2. The planted bi-section model, posteriors and tests

In this section, we introduce the model and prepare the theorems on exact and almost-exact community detection in the next section. We consider prior and posterior, define metrics for community assignments, we derive posterior concentration based on test functions and we prove the existence of suitable test functions.

2.1. The planted bi-section model

In a stochastic block model, each vertex is assigned to one of $K \geq 2$ communities through an unobserved *community assignment vector* θ'_n. Each vertex belongs to a community and any edge occurs (independently of others) with a probability depending on the communities of the vertices that it connects. In the *planted bi-section model*, there are only two communities ($K = 2$) and, at the nth iteration ($n \geq 1$), there are $2n$ vertices (labelled with indices $1 \leq i \leq 2n$), n in each community, with community assignment vector $\theta'_n \in \Theta'_n$ (with components $\theta'_1, \ldots, \theta'_{2n} \in \{0, 1\}$), where Θ'_n is the subset of $\{0, 1\}^{2n}$ of all finite binary sequences that contain as many ones as zeros. Denote that space in which the random graph X^n takes its values by $\mathscr{X}_n$ (*e.g.*, represented by its adjacency matrix with entries $\{X_{ij} : 1 \leq i, j \leq 2n\}$) and its distribution by $P_{\theta'_n,n}$. The (n-dependent) probability of an edge occurring ($X_{ij} = 1$) between vertices $1 \leq i, j \leq 2n$ *within the same community* is denoted $p_n \in (0, 1)$; the probability of an edge *between communities* is denoted $q_n \in (0, 1)$,

$$Q_n(\theta'_{n,i}, \theta'_{n,j}) := P_{\theta'_n,n}(X_{ij} = 1) = \begin{cases} p_n, & \text{if } \theta'_{n,i} = \theta'_{n,j}, \\ q_n, & \text{if } \theta'_{n,i} \neq \theta'_{n,j}. \end{cases} \tag{3}$$

Note that if $p_n = q_n$, X^n is the Erdős-Rényi graph $G(2n, p_n)$ and the community assignment $\theta'_n \in \Theta'_n$ is not identifiable. Another identifiability issue that arises is that the model is invariant under interchange of community labels 0 and 1. This is expressed in the parameter spaces Θ'_n through equivalence relations: $\theta'_{1,n} \sim_n \theta'_{2,n}$, if $\theta'_{2,n} = \neg\theta'_{1,n}$ (by component-wise negation). To prevent non-identifiability, we parametrize the model for X^n in terms of a parameter θ_n in a quotient space $\Theta_n = \Theta'_n/\sim_n$, for every $n \geq 1$. For $\theta'_n \in \Theta'_n$ we denote the equivalence class $\{\theta'_n, \neg\theta'_n\}$ by θ_n. Note that the set Θ_n can be identified with the set of partitions of $\{1, \ldots, 2n\}$ consisting of exactly two sets with n elements, via the identification

$$\theta_n \longleftrightarrow \left\{ \left\{ i : \theta'_{n,i} = 0 \right\}, \left\{ i : \theta'_{n,i} = 1 \right\} \right\},$$

Note that this is independent of the choice of the representation and that Q_n is well-defined on $\Theta_n \times \Theta_n$.

Given true parameters $\theta_{0,n} \in \Theta_n$ ($n \geq 1$), choose representations $\theta'_{0,n} \in \Theta'_n$ and define $Z_n(\theta'_0) \subset \{1, \ldots, 2n\}$ to be *community zero* (the set of all those i such that $\theta'_{0,i} = 0$) and call the complement $Z^c_n(\theta'_0)$ *community one*. For the questions concerning exact recovery and detection, we are interested in the sets $V'_{n,k} \subset \Theta'_n$, defined to contain all those θ'_n that differ from $\theta'_{0,n}$ by exactly k exchanges of pairs: for $\theta'_n \in \Theta'_n$ we have $\theta'_n \in V'_{n,k}$, if the set of vertices in community zero *cf.* $\theta'_{0,n}$, $Z(\theta'_{0,n}) = \{1 \leq i \leq 2n : \theta'_{0,n,i} = 0\}$, from which we leave out the set of vertices in community zero *cf.* θ'_n, $Z(\theta'_n) = \{1 \leq i \leq 2n : \theta'_{n,i} = 0\}$, has k elements. Conversely, for any $\theta'_{1,n}$ and $\theta'_{2,n}$ in Θ'_n, we denote the minimal number of pair-exchanges necessary to take $\theta'_{1,n}$ into $\theta'_{2,n}$ by $k'(\theta'_{1,n}, \theta'_{2,n})$. Note that $k'(\theta'_{1,n}, \neg\theta'_{2,n}) = n - k'(\theta'_{1,n}, \theta'_{2,n})$, which leads to a metric on Θ_n,

$$k(\theta_{1,n}, \theta_{2,n}) = k'(\theta'_{1,n}, \theta'_{2,n}) \wedge k'(\theta'_{1,n}, \neg\theta'_{2,n}). \tag{4}$$

Note that k is independent of choice of the representations and that k takes values in $\{0, \ldots, \lfloor n/2 \rfloor\}$. Now define,

$$V_{n,k} = V_{n,k}(\theta_{0,n}) = \left\{ \theta_n : k(\theta_n, \theta_{0,n}) = k \right\} = \left\{ \theta_n : \theta'_n \in V'_{n,k} \right\}, \tag{5}$$

for $k \in \{1, \ldots, \lfloor n/2 \rfloor\}$. Given some sequence (k_n) of positive integers we then define V_n as the disjoint union,

$$V_n = \bigcup_{k=k_n}^{\lfloor n/2 \rfloor} V_{n,k}. \tag{6}$$

Since we can choose two subsets of k elements from two sets of size n in $\binom{n}{k}^2$ ways, the cardinality of $V_{n,k}$ is $\binom{n}{k}^2$, when $k < n/2$ and $\frac{1}{2}\binom{n}{n/2}^2$ when n is even and $k = n/2$. In both cases the number of elements in $V_{n,k}$ is therefore bounded by $\binom{n}{k}^2$.

With that perspective on the parameter in mind, we note that the likelihood with observed graph X^n is given by,

$$p_{\theta,n}(X^n) = \prod_{i<j} Q_n(\theta_{n,i}, \theta_{n,j})^{X_{ij}} (1 - Q_n(\theta_{n,i}, \theta_{n,j}))^{1-X_{ij}}. \tag{7}$$

For the *sparse versions* of the planted bi-section model, we also define edge probabilities that vanish with growing n: take (a_n) and (b_n) such that $a_n \log n = np_n$ and $b_n \log n = nq_n$ for the so-called *Chernoff–Hellinger phase*; take (c_n) and (d_n) such that $c_n = np_n$ and $d_n = nq_n$ for the so-called *Kesten-Stigum phase*. The fact that we do not allow loops (edges that connect vertices with themselves) leaves room for $2 \cdot \frac{1}{2}n(n-1) + n^2 = 2n^2 - n = \frac{1}{2} \cdot (2n)(2n-1)$ possible edges in the random graph X^n observed at iteration n.

Our first statistical question of interest is reconstruction of the (frequentist) true community assignment vectors θ_n *consistently*, that is, (close to) correctly with probability growing to one as n tends to infinity. Consistency can be formulated in various ways, and we consider two of those formulations below.

Definition 2.1. Let $\theta_{0,n} \in \Theta_n$ be given. An estimator sequence $\hat{\theta}_n : \mathscr{X}_n \to \Theta_n$ is said to *recover the community assignment* $\theta_{0,n}$ *exactly* if,

$$P_{\theta_{0,n}}\big(\hat{\theta}_n(X^n) = \theta_{0,n}\big) \to 1,$$

as n tends to infinity, that is, if $\hat{\theta}_n$ coincides with the correct community assignment vector with high probability.

We also relax this consistency requirement somewhat in the form of the following definition, *cf.* Mossel et al. (2016) and others: for $n \geq 1$ and two community assignments $\theta_{0,n}, \theta_n \in \Theta_n$, let $k(\theta_n, \theta_{0,n})$ denote the *minimal number of pair exchanges needed to transform* θ_n *into* $\theta_{0,n}$ (for further details, see the definition of k, just before Eq. (6) below).

Definition 2.2. Let $\theta_{0,n} \in \Theta_n$ be given. An estimator sequence $\hat{\theta}_n : \mathscr{X}_n \to \Theta_n$ is said to *recover* $\theta_{0,n}$ *almost-exactly*, if $k(\hat{\theta}_n, \theta_{0,n})$ is of order $o_P(n)$ under $P_{\theta_{0,n}}$. If, for some sequence $l_n = o(n)$,

$$P_{\theta_{0,n}}\big(k(\hat{\theta}_n, \theta_{0,n}) \leq l_n\big) \to 1,$$

as n tends to infinity, we say that $\hat{\theta}_n$ *recovers* $\theta_{0,n}$ *with error rate* l_n.

The second statistical problem we study is posterior-based, asymptotic, frequentist uncertainty quantification for the community assignment vector. We recall the central definition and its asymptotic version for later reference.

Definition 2.3. For fixed $n \geq 1$ and some $0 < a < 1$, a set-valued map $x^n \mapsto C(x^n)$ defined on $\mathscr{X}_n$ such that, for all $\theta_n \in \Theta_n$, $P_{\theta_n,n}(\theta_n \in C(X^n)) \geq 1 - a$, is called a confidence set of level $1 - a$. If the levels $1 - a_n$ of n-dependent confidence sets $C_n(X^n)$ go to 1 as n tends to infinity, the $C_n(X^n)$ are said to be asymptotically consistent.

Bayesian notions of uncertainty quantification, in particular credible sets, and the way in which they are enlarged to form confidence sets, is discussed in Appendix B.

2.2. Prior, posterior and test functions

Consider the sequence of experiments in which we observe random graphs $X^n \in \mathscr{X}_n$ generated by the planted bi-section model of definition (3). In much of the literature on the stochastic block model, the Bayesian approach is chosen: we pick prior distributions π_n for all Θ_n, $(n \geq 1)$ and calculate the posterior distribution: denoting the likelihood by $p_{\theta,n}(X^n)$, the posterior for the parameter θ_n is written as a fraction of sums, for all $A \subset \Theta_n$,

$$\Pi(A|X^n) = \sum_{\theta_n \in A} p_{\theta,n}(X^n)\,\pi_n(\theta_n) \Bigg/ \sum_{\theta_n \in \Theta_n} p_{\theta,n}(X^n)\,\pi_n(\theta_n),$$

where $\pi_n : \Theta_n \to [0,1]$ is the probability mass function for the prior Π_n. Below we only consider *uniform priors* (Π_n) for $\theta_n \in \Theta_n$, so for all $n \geq 1$ and $\theta_n \in \Theta_n$, $\pi(\theta_n) = \pi_n := (|\Theta_n|)^{-1}$.

Remark 2.4. To motivate our choice for a *uniform prior*, note that for community detection in the planted bi-section model, all values of the community assignment are equivalent, in the sense that the statistical problem is invariant under permutation of the vertices. So, non-uniformity of the prior would imply a strictly subjective bias, which has no place in the application of posteriors for *frequentist* inference. Additionally, non-uniformity of the prior, although it would raise posterior concentration of mass at particular values of the community assignment vector, would also go at the expense of posterior mass at certain other values. Our limits for posterior concentration are formulated *for all values of the community assignment*, so non-uniformity can only make assertions weaker and conditions stronger.

According to lemma 2.2 in Kleijn (2021) (with $B_n = \{\theta_{0,n}\}$), for any measurable sequence $\phi_{k,n} : \mathscr{X}_n \to [0,1]$ $(k \geq 1, n \geq 1)$ (following Le Cam, 1986, we refer to such functions as *test functions* in what follows), we have,

$$\begin{aligned} P_{\theta_0,n}\Pi(V_n|X^n) &= \sum_{k=k_n}^{\lfloor n/2 \rfloor} P_{\theta_0,n}\Pi(V_{n,k}|X^n) \\ &\leq \sum_{k=k_n}^{\lfloor n/2 \rfloor} \Big(P_{\theta_0,n}\phi_{k,n}(X^n) + \sum_{\theta_n \in V_{n,k}} P_{\theta_n,n}(1 - \phi_{k,n}(X^n)), \Big) \end{aligned}$$

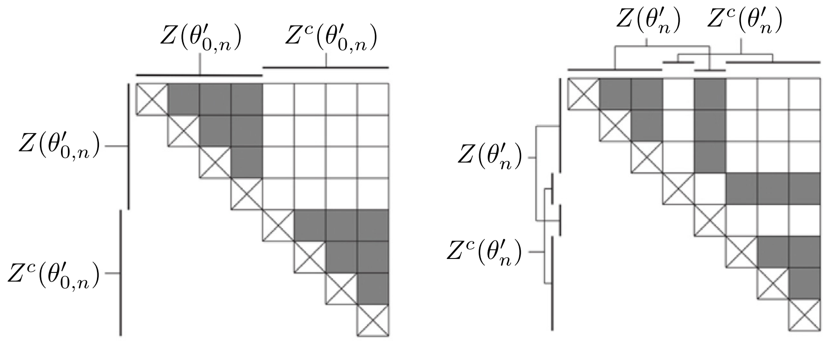

Fig. 1. Community assignments and edge probabilities according to $\theta'_{0,n}$ and to θ'_n for $n = 4$ and $k = 1$. Vertex sets $Z(\cdot)$ and $Z^c(\cdot)$ correspond to communities zero and one for the given community assignment. Dark squares correspond to edges that occur with (within-community) probability p_n, and light squares to edges that occur with (between-community) probability q_n.

for every $n \geq 1$. Suppose that for any $k \geq 1$ there exists a sequence $(a_{n,k})_{n\geq 1}$, $a_{n,k} \downarrow 0$ and, for any $\theta_n \in V_{n,k}$, a test function $\phi_{\theta_n,n}$ that distinguishes $\theta_{0,n}$ from θ_n as follows,

$$P_{\theta_0,n}\phi_{\theta_n,n}(X^n) + P_{\theta_n,n}(1 - \phi_{\theta_n,n}(X^n)) \leq a_{n,k}, \tag{8}$$

for all $n \geq 1$. Then using test functions $\phi_{k,n}(X^n) = \max\{\phi_{\theta_n,n}(X^n) : \theta_n \in V_{n,k}\}$, as well as the fact that,

$$P_{\theta_0,n}\phi_{k,n}(X^n) \leq \sum_{\theta_n \in V_{n,k}} P_{\theta_0,n}\phi_{\theta_n,n}(X^n),$$

we see that,

$$\begin{aligned} P_{\theta_0,n}\Pi(V_n|X^n) &\leq \sum_{k=k_n}^{\lfloor n/2 \rfloor} \sum_{\theta_n \in V_{n,k}} \Big(P_{\theta_{0,n},n}\phi_{\theta_n,n}(X^n) + P_{\theta_n,n}(1 - \phi_{\theta_n,n}(X^n))\Big) \\ &\leq \sum_{k=k_n}^{\lfloor n/2 \rfloor} \binom{n}{k}^2 a_{k,n}. \end{aligned} \tag{9}$$

This inequality forms the basis for the results in the next section on exact recovery and almost-exact recovery.

2.3. Existence of suitable test functions

Given $n \geq 1$ and two community assignment vectors $\theta_{0,n}, \theta_n \in \Theta_n$, we are interested in calculation of the likelihood ratio $dP_{\theta,n}/dP_{\theta_0,n}$, because it determines testing power as well as the various forms of remote contiguity that play a role.

Choose representations θ'_0 of θ_0 and θ' of θ so that $k'(\theta'_0, \theta') = k(\theta_0, \theta)$, where k and k' are as in Section 3. Recall that, $Z_n(\theta'_0) \subset \{1, \ldots, 2n\}$ is community zero and the complement $Z_n^c(\theta'_0)$ community one. For the sake of presentation (in Fig. 1), re-label the vertices such that $Z(\theta'_0) = \{1, \ldots, n\}$ and $Z^c(\theta'_0) = \{n+1, \ldots, 2n\}$. In the case $n = 4$, Fig. 1 shows edge probabilities in the familiar block arrangement.

The likelihood under θ_0 is given by Eq. (7), with $\theta = \theta_0$. If we assume that $\theta'_{0,n}$ and θ'_n differ by k pair-exchanges among respective members of communities zero and one, then a look at Fig. 1 reveals that the likelihood-ratio depends only on the edges for which exactly one of its end-points changes community. Define,

$$\begin{aligned} A_n &= \{(i,j) \in \{1, \ldots, 2n\}^2 : i < j,\ \theta'_{0,n,i} = \theta'_{0,n,j},\ \theta'_{n,i} \neq \theta'_{n,j}\}, \\ B_n &= \{(i,j) \in \{1, \ldots, 2n\}^2 : i < j,\ \theta'_{0,n,i} \neq \theta'_{0,n,j},\ \theta'_{n,i} = \theta'_{n,j}\}. \end{aligned}$$

Also define,

$$(S_n, T_n) := \Big(\sum\{X_{ij} : (i,j) \in A_n\}, \sum\{X_{ij} : (i,j) \in B_n\}\Big),$$

and note that the likelihood ratio can be written as,

$$\frac{p_{\theta,n}}{p_{\theta_0,n}}(X^n) = \left(\frac{1-p_n}{p_n}\,\frac{q_n}{1-q_n}\right)^{S_n - T_n}, \tag{10}$$

where,

$$(S_n, T_n) \sim \begin{cases} \mathrm{Bin}(2k(n-k), p_n) \times \mathrm{Bin}(2k(n-k), q_n), & \text{if } X^n \sim P_{\theta_0,n}, \\ \mathrm{Bin}(2k(n-k), q_n) \times \mathrm{Bin}(2k(n-k), p_n), & \text{if } X^n \sim P_{\theta,n}. \end{cases} \tag{11}$$

Based on that, we derive the following lemma.

Lemma 2.5. *Let $n \geq 1$, $\theta_{0,n}, \theta_n \in \Theta_n$ be given. Assume that $\theta_{0,n}$ and θ_n differ by k pair-exchanges. Then there exists a test function $\phi_n : \mathscr{X}_n \to [0,1]$ such that,*

$$P_{\theta_0,n}\phi_n(X^n) + P_{\theta,n}(1-\phi_n(X^n)) \leq a_{n,k},$$

with testing power,

$$a_{n,k} = \left(1 - p_n - q_n + 2p_n q_n + 2\sqrt{p_n(1-p_n)}\sqrt{q_n(1-q_n)}\right)^{2k(n-k)}.$$

Proof. The likelihood ratio test $\phi_n(X^n)$ has testing power bounded by the so-called Hellinger transform,

$$P_{\theta_0,n}\phi_n(X^n) + P_{\theta,n}(1-\phi_n(X^n)) \leq \inf_{0\leq\alpha\leq1} P_{\theta_0,n}\left(\frac{p_{\theta,n}}{p_{\theta_0,n}}(X^n)\right)^{\alpha},$$

(see, *e.g.*, Le Cam (1986) and proposition 2.6 in Kleijn (2021)). Using $\alpha = 1/2$ (which is the minimum) and the independence of S and T, we find that,

$$P_{\theta_0,n}\left(\frac{p_{\theta,n}}{p_{\theta_0,n}}(X^n)\right)^{\frac{1}{2}} = P_{\theta_0,n}\left(\frac{p_n}{1-p_n}\,\frac{1-q_n}{q_n}\right)^{\frac{1}{2}(T_n-S_n)} = Pe^{\frac{1}{2}\lambda_n S_n}\,Pe^{-\frac{1}{2}\lambda_n T_n},$$

where $\lambda_n := \log(1-p_n) - \log(p_n) + \log(q_n) - \log(1-q_n)$ and (S_n, T_n) are distributed binomially, as in the first case of (11). Using the moment-generating function of the binomial distribution, we conclude that,

$$\begin{aligned}
&P_{\theta_0,n}\left(\frac{p_{\theta,n}}{p_{\theta_0,n}}(X^n)\right)^{1/2}\\
&\quad= \left(\left(1 - p_n + p_n\left(\frac{1-p_n}{p_n}\,\frac{q_n}{1-q_n}\right)^{1/2}\right)\right.\\
&\qquad\qquad \left.\times\left(1 - q_n + q_n\left(\frac{p_n}{1-p_n}\,\frac{1-q_n}{q_n}\right)^{1/2}\right)\right)^{2k(n-k)}\\
&\quad= \left(\left((1-p_n) + p_n^{1/2}q_n^{1/2}\left(\frac{1-p_n}{1-q_n}\right)^{1/2}\right)\right.\\
&\qquad\qquad \left.\times\left((1-q_n) + p_n^{1/2}q_n^{1/2}\left(\frac{1-q_n}{1-p_n}\right)^{1/2}\right)\right)^{2k(n-k)}\\
&\quad= \left((1-p_n)(1-q_n) + 2\big(p_nq_n(1-p_n)(1-q_n)\big)^{1/2} + p_nq_n\right)^{2k(n-k)},
\end{aligned}$$

which proves the assertion. □

3. Exact and almost-exact posterior recovery of communities

In this section, we combine inequality (9) with the test functions of Section 2.3 to arrive at two posterior concentration results, for exact and almost-exact recovery of the community structure.

3.1. Posterior consistency: exact recovery

For exact recovery, we are interested in the expected posterior masses of subsets of Θ_n of the form:

$$V_n = \{\theta_n \in \Theta_n : \theta_n \neq \theta_{0,n}\} = \bigcup_{k=1}^{\lfloor n/2\rfloor} V_{n,k}.$$

The theorem states a sufficient condition for (p_n) and (q_n), which is related to requirement (2) in the Chernoff–Hellinger phase.

Theorem 3.1. *For some $\theta_{0,n} \in \Theta_n$, assume that $X^n \sim P_{\theta_0,n}$, for every $n \geq 1$. If we equip every Θ_n with its uniform prior and $p_n = a_n n^{-1}\log n$ and $q_n = b_n n^{-1}\log n$ are of order $O(n^{-1/2})$, with (a_n), (b_n) such that,*

$$(a_n + b_n - 2\sqrt{a_nb_n} - 1)\log n \to \infty, \tag{12}$$

then,

$$\Pi\big(\theta_n = \theta_{0,n} \bigm| X^n\big) \xrightarrow{P_{\theta_{0,n}}} 1, \tag{13}$$

as n tends to infinity, i.e., the posterior recovers the community assignment exactly.

Proof. We first give an alternative formulation of condition (13) that is more suitable for the proof. Let $\delta_n = a_n + b_n - 2\sqrt{a_n b_n} - 1 = (\sqrt{a_n} - \sqrt{b_n})^2 - 1$, and (13) implies that $\delta_n \log n \to \infty$. In particular, for large enough n, $0 < \delta_n < 1$. Conversely, assume that, there is a sequence δ_n such that for n sufficiently large, $0 < \delta_n < 1$, $\delta_n \log n \to \infty$ as n tends to infinity, and,

$$(\sqrt{a_n} - \sqrt{b_n})^2 \geq 1 + \delta_n, \tag{14}$$

for n sufficiently large. Conclude that (12) is equivalent to (14).

According to Lemma 2.5, for every $n \geq 1$, $k \geq 1$ and given $\theta_{0,n}$, there exists a test sequence satisfying (8) with $a_{n,k} = (1-\mu_n)^{2k(n-k)}$, where,

$$\mu_n = p_n + q_n - 2p_n q_n - 2(p_n(1-p_n)q_n(1-q_n))^{1/2},$$

in $[0, 1]$. Start by noting that $\mu_n \geq p_n + q_n - 2p_n q_n - 2\sqrt{p_n}\sqrt{q_n} = (\sqrt{p_n} - \sqrt{q_n})^2 - 2p_n q_n$, so that assumption (14) implies that,

$$1 - \mu_n \leq 1 - (1+\delta_n)\frac{\log n}{n} + \frac{2a_n b_n (\log n)^2}{n^2}. \tag{15}$$

It follows from the assumption that p_n and q_n are of order $O(n^{-1/2})$ and $\delta_n \log n \to \infty$, that that last term on the right-hand side is smaller than $\frac{1}{2}\delta_n \log n$ for large enough n. Therefore, $1 - \mu_n \leq 1 - (1 + \delta_n/2)n^{-1}\log n$, and Lemma A.3 says that,

$$(1-\mu_n)^n \leq e^{-(1+\delta_n/2)\log n}. \tag{16}$$

Using that $\binom{n}{k}^2 \leq \binom{2n}{2k}$ in (9), we find,

$$\begin{aligned}
&P_{\theta_{0,n}}\Pi(V_n|X^n)\\
&\quad\leq \sum_{k=1}^{\lfloor n/2\rfloor}\binom{n}{k}^2(1-\mu_n)^{2k(n-k)} \leq \sum_{k=1}^{\lfloor n/2\rfloor}\binom{2n}{2k}(1-\mu_n)^{2k(n-k)}\\
&\quad\leq \sum_{\ell=2}^{2\lfloor n/2\rfloor}\binom{2n}{\ell}(1-\mu_n)^{\ell(n-\ell/2)} \leq \sum_{k=1}^{n}\binom{2n}{k}(1-\mu_n)^{k(n-k/2)}\\
&\quad\leq \sum_{k=1}^{\lfloor \delta_n n/2\rfloor}\binom{2n}{k}(1-\mu_n)^{k(n-k/2)} + \sum_{\lceil \delta_n n/2\rceil}^{n}\binom{2n}{k}(1-\mu_n)^{k(n-k/2)}.
\end{aligned} \tag{17}$$

We will analyse the two sums on the right-hand side separately. Regarding the first sum, note that $n - k/2 \geq (1-\delta_n/4)n$ for all integers $1 \leq k \leq \delta_n n/2$, so,

$$\begin{aligned}
&\sum_{k=1}^{\lfloor \delta_n n/2\rfloor}\binom{2n}{k}(1-\mu_n)^{k(n-k/2)} \leq \sum_{k=1}^{2n}\binom{2n}{k}(1-\mu_n)^{k(1-\delta_n/4)n}\\
&\qquad\leq \sum_{k=1}^{2n}\binom{2n}{k}e^{-k(1-\delta_n/4)(1+\delta_n/2)\log n} \leq \sum_{k=1}^{2n}\binom{2n}{k}e^{-k(1+\delta_n/8)\log n},
\end{aligned} \tag{18}$$

where we use that $(1-\delta_n/4)(1+\delta_n/2) \geq 1 + \delta_n/8$. By the binomial theorem and Lemma A.3 the first sum in (17) is bounded by,

$$\left(1 + e^{-(1+\delta_n/8)\log n}\right)^{2n} - 1 \leq e^{2ne^{-(1+\delta_n/8)\log n}} - 1 = e^{2e^{-(\delta_n/8)\log n}} - 1 \to 0, \tag{19}$$

as n tends to infinity.

Regarding the second sum on the right-hand side of (17), it is noted that $x(1-x)$ attains its minimum at $x = a$ on the interval $[a, 1/2], 0 < a < 1/2$, so that $k(n-k/2) = 2n^2(k/(2n))(1-k/(2n)) \geq \delta_n(1-\delta_n/4)n^2/2$, for all integers $\delta_n n/2 \leq k \leq n$. Therefore,

$$(1-\mu_n)^{k(n-k/2)} \leq (1-\mu_n)^{\delta_n(1-\delta_n/4)n^2/2} \leq e^{-\frac{1}{2}n\delta_n(1-\delta_n/4)(1+\delta_n/2)\log n}. \tag{20}$$

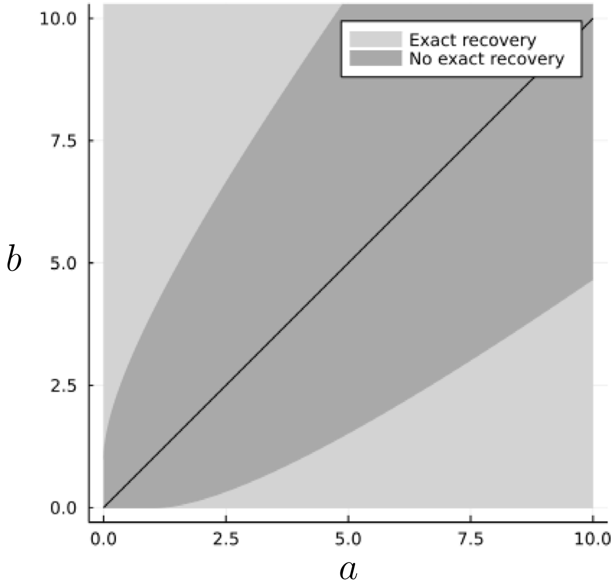

Fig. 2. The phase diagram for the sparse planted bi-section model in the Chernoff–Hellinger phase: values of (a, b) that are too close to the diagonal, resemble the Erdős-Rényi model (in which there are no communities) too closely, rendering the community structure non-exactly-recoverable.

Substituting, we find,

$$\begin{aligned}\sum_{\lceil \delta_n n/2 \rceil}^{2n} \binom{2n}{k} (1-\mu_n)^{k(n-k/2)} &\le e^{-\frac{1}{2} n\delta_n (1-\delta_n/4)(1+\delta_n/2)\log n} \sum_{k=0}^{2n} \binom{2n}{k} \\ &= e^{2n\log 2 - \frac{1}{2} n\delta_n(1-\delta_n/4)(1+\delta_n/2)\log n} \to 0,\end{aligned} \tag{21}$$

as n tends to infinity. The latter limit and (19) prove the assertion. □

Up to the $\frac{1}{2}\log\log(n)$-term, condition (12) is equal to (the necessary and sufficient) condition (2) of Mossel et al. (2016). In fact there is a trade-off: (12) is slightly weaker than (2), but (2) applies only if there exists a $C > 0$ such that $C^{-1} \le a_n, b_n \le C$ for large enough n (Mossel et al., 2016; Zhang and Zhou, 2016). This bound excludes some interesting examples in which one of the sequences (a_n) and (b_n) may fade away with growing n or equal zero outright. For instance, if $b_n = 0$ and $\liminf_n a_n > 1$, edges between communities are completely absent but, separately, the Erdős-Rényi graphs spanned by vertices in $Z_n(\theta_0')$ and $Z_n^c(\theta_0')$ respectively are connected with high probability. Similarly, if $a_n = 0$ and $\liminf_n b_n > 1$, the posterior succeeds in exact recovery: possibly with b_n above 1, edges between communities are abundant enough to guarantee the existence of a path in X^n that visits all vertices at least once, with high probability. It is tempting to state the following, well-known (Abbe et al., 2016; Abbe, 2018; Mossel et al., 2016) sufficient condition for the sequences $a_n > 0$ and $b_n > 0$:

$$(\sqrt{a_n} - \sqrt{b_n})^2 > c, \text{ for some } c > 1 \text{ and } n \text{ large enough}, \tag{22}$$

(even though it ignores the logarithm in (2)). Fig. 2 provides a ‘phase diagram’ delineating the choices for (a, b) such that $a_n = a$ and $b_n = b$ leads to exact recovery (analogous to Abbe (2018, theorem 3)).

Corollary 3.2. *Under the conditions of Theorem 3.1, the MAP-/ML-estimator recovers $\theta_{0,n}$ exactly.*

Proof. Due to the uniformity of the prior, for every $n \ge 1$, maximization of the posterior density (with respect to the counting measure) on Θ_n, is the same as maximization of the likelihood. Due to (13), the posterior densities in the points $\theta_{0,n}$ in Θ_n converge to one in $P_{\theta_{0,n}}$-probability. Accordingly, the point of maximization is $\theta_{0,n}$ with high probability. □

3.2. Posterior consistency: almost-exact recovery

For the case of almost-exact recovery, the requirement of convergence is less stringent: as said, Mossel et al. (2016, proposition 2.9) states that condition (1) is *necessary and sufficient* for almost-exact recovery. Below we show that posteriors with uniform priors *recover* the true community assignment *almost exactly* if (1) holds.

We are interested in the expected posterior masses of subsets of Θ_n of the form:

$$W_n = \bigcup_{k=k_n}^{\lfloor n/2 \rfloor} V_{n,k},$$

for a sequence k_n of order $o(n)$ or $O(n)$: the posterior concentrates on community assignments θ_n that differ from $\theta_{0,n}$ by no more than k_n pair exchanges.

Theorem 3.3. *For some $\theta_{0,n} \in \Theta_n$, let $X^n \sim P_{\theta_0,n}$ for every $n \geq 1$. If we equip all Θ_n with uniform priors and edge-probabilities (p_n), (q_n) and error rates (k_n) are such that,*

$$\frac{n}{k_n}\Big(1 - p_n - q_n + 2p_n\, q_n + 2\sqrt{p_n(1-p_n)q_n(1-q_n)}\Big)^{n/2} \to 0, \tag{23}$$

as n tends to infinity, then,

$$\Pi(W_n|X^n) \xrightarrow{P_0} 0, \tag{24}$$

as n tends to infinity, i.e., the posterior recovers $\theta_{0,n}$ with error rate k_n.

Proof. According to Lemma 2.5, for every $n \geq 1$, $k \geq 1$ and given $\theta_{0,n}$, there exists a test sequence satisfying (8) with $a_{n,k} = (1-\mu_n)^{2k(n-k)}$. Therefore, using the inequalities $\binom{2n}{k} \leq \frac{(2n)^k}{k!}$ and $(n+m)! \geq n!m!$, the Stirling lower bound formula, and finally our assumption $n(1-\mu_n)^{n/2}/k_n \to 0$ (μ_n as defined in the proof of Theorem 3.1, we see that for big enough n,

$$\begin{aligned}
P_{\theta_0,n}\Pi(W_n|X^n) &\leq \sum_{k=k_n}^{\lfloor n/2\rfloor} \binom{n}{k}^2 (1-\mu_n)^{2k(n-k)} \\
&\leq \sum_{k=2k_n}^{n} \binom{2n}{k}(1-\mu_n)^{k(n-k/2)} \leq \sum_{k=2k_n}^{\infty} \frac{1}{k!}(2n)^k(1-\mu_n)^{kn/2} \\
&\leq \frac{\big(2n(1-\mu_n)^{n/2}\big)^{2k_n}}{(2k_n)!} e^{2n(1-\mu_n)^{n/2}}.
\end{aligned}$$

It then follows that,

$$\begin{aligned}
P_{\theta_0,n}\Pi(W_n|X^n) &\leq \frac{1}{\sqrt{4\pi k_n}}\left(\frac{n(1-\mu_n)^{n/2}}{k_n}\right)^{2k_n} e^{2k_n + 2n(1-\mu_n)^{n/2}} \\
&\leq \frac{1}{\sqrt{4\pi k_n}}\left(\frac{n(1-\mu_n)^{n/2}}{k_n} e^{1+n(1-\mu_n)^{n/2}/k_n}\right)^{2k_n} \\
&\leq \frac{n(1-\mu_n)^{n/2}}{k_n} e^{1+n(1-\mu_n)^{n/2}/k_n},
\end{aligned}$$

which converges to zero as $n \to \infty$. $\square$

Example 3.4. Note that if $p_n, q_n = O(n^{-1}) = o(1)$, we may expand,

$$\sqrt{p_n} - \sqrt{q_n} = \frac{1}{2\sqrt{\frac{1}{2}(p_n+q_n)}}(p_n - q_n) + O(|p_n - q_n|^2),$$

which means that,

$$\mu_n = (\sqrt{p_n} - \sqrt{q_n})^2 + O(n^{-2}) = \frac{(p_n-q_n)^2}{2(p_n+q_n)} + O(n^{-2}).$$

Assuming only that $n(p_n - q_n)^2 > 2(p_n + q_n)$, as in Decelle et al. (2011b,a), we would arrive at the conclusion that $n\mu_n > 1 + O(n^{-1})$, which is insufficient in the proof of Theorem 3.3. Note that a non-divergent choice $k_n = O(1)$ forces us back into the Chernoff–Hellinger phase where exact recovery is possible.

Corollary 3.5. *Under the conditions of Theorem 3.3 with (p_n) and (q_n) such that,*

$$n\Big(p_n + q_n - 2p_n\, q_n - 2\sqrt{p_n(1-p_n)q_n(1-q_n)}\Big) \to \infty, \tag{25}$$

as n tends to infinity, posteriors recover $\theta_{0,n}$ partially,

$$\Pi\Big(\, k(\theta_n, \theta_{0,n}) \ge \beta n \;\Big|\; X^n \,\Big) \xrightarrow{P_0} 0,$$

for any *fraction $\beta \in (0, \frac{1}{2})$, which implies that the posterior recovers $\theta_{0,n}$ almost-exactly.*

Proof. Let $\beta \in (0, \frac{1}{2})$ be given. Follow the proof of Theorem 3.3 with $k_n = \beta n$ and note that,

$$P_{\theta_{0,n}}\Pi\Big(\, k(\theta_n, \theta_{0,n}) \ge \beta n \;\Big|\; X^n \,\Big) \le \frac{1}{\beta}(1-\mu_n)^{n/2} e^{1+\beta^{-1}(1-\mu_n)^{n/2}}.$$

Due to Eq. (25),

$$(1-\mu_n)^{n/2} = \Big(1 - p_n - q_n + 2p_n q_n + 2\sqrt{p_n(1-p_n)q_n(1-q_n)}\Big)^{n/2} \to 0,$$

so $P_{\theta_{0,n}}\Pi(k(\theta_n, \theta_{0,n}) \ge \beta n \mid X^n) \to 0$. For almost-exact recovery, let $\beta_m \downarrow 0$ be given; if we let $m(n)$ go to infinity slowly enough, posterior convergence continues to hold with β equal to $\beta_{m(n)}$. $\square$

Condition (25) says that $n\mu_n \to \infty$ is *sufficient* for almost-exact posterior recovery; but as shown in Mossel et al. (2016, proposition 2.10), it is *also necessary* for any form of almost-exact recovery. We conclude that if there exist any estimators $\hat{\theta}_n$ that recover the community assignment almost exactly, then posteriors with uniform priors also recover the community assignment almost-exactly.

4. Uncertainty quantification for community structure

Our first results on uncertainty quantification are obtained with the help of the results in the previous section: if we know that the sequences (p_n) and (q_n) satisfy requirements like (12) or (23), so that exact or almost-exact recovery is guaranteed, then a consistent sequence of confidence sets is easily constructed from credible sets, as shown in Section 4.1 and the sizes of these credible sets as well as the sizes of associated confidence sets are controlled. If we cannot guarantee (12) or (23), or if we require explicit control over confidence levels, confidence sets can still be constructed from credible sets under conditions requiring that credible levels grow to one quickly enough. Enlargement of credible sets may be used to mitigate this condition, whenever we are close to the Erdős-Rényi sub-model, as discussed in Section 4.2.

Regarding the sizes of credible sets, the most natural way to compile a minimal-order credible set $E_n(X^n)$ in a discrete space like Θ_n, is to calculate the posterior weights $\Pi(\{\theta_n\}|X^n)$ of all $\theta_n \in \Theta_n$, order Θ_n by decreasing posterior weight into a finite sequence $\{\theta_{n,1}, \theta_{n,2}, \ldots, \theta_{n,|\Theta_n|}\}$ and define $E_n(X^n) = \{\theta_{n,1}, \ldots, \theta_{n,m}\}$, for the smallest $m \ge 1$ such that $\Pi(E_n(X^n)|X^n)$ is greater than or equal to the required credible level. To provide guarantees regarding the sizes of credible sets, one would like to show that these $E_n(X^n)$ are of an order that is upper bounded with high probability. (Although it is not so clear what the upper bound should be, ideally.)

Here we shall follow a different path based on the smallest number $k(\theta_n, \eta_n)$ of pair-exchanges between (two representations θ'_n and η'_n in Θ'_n of) θ_n and η_n respectively, see (4). The map $k : \Theta_n \times \Theta_n \to \{0, 1, \ldots, \lfloor n/2 \rfloor\}$ is interpreted in a role similar to that of a metric on larger parameter spaces: the *diameter* $\mathrm{diam}_n(C)$ of a subset $C \subset \Theta_n$ is,

$$\mathrm{diam}_n(C) = \max\Big\{k(\theta_n, \eta_n) : \theta_n, \eta_n \in C\Big\}.$$

by definition.

4.1. Posterior recovery and confidence sets

If the posteriors concentrate amounts of mass on $\{\theta_{0,n}\}$ arbitrarily close to one with growing n, then a sequence of credible sets of a certain fixed level contains $\theta_{0,n}$ for large enough n. If such posterior concentration occurs with high $P_{\theta_{0,n}}$-probability, then the sequence of credible sets is also an asymptotically consistent sequence of confidence sets. We formalize and prove this observation in the following theorem. (A real-valued sequence (c_n) is said to be bounded away from zero in the limit, if $\liminf_n c_n > 0$).

Theorem 4.1. *Let (c_n) be bounded away from zero in the limit. Suppose that the posterior recovers the communities exactly. Then any sequence (D_n) of (P_n^{Π}-almost-sure) credible sets of levels c_n satisfies,*

$$P_{\theta_{0,n}}\Big(\, \theta_{0,n} \in D_n(X^n) \,\Big) \to 1,$$

i.e., (D_n) is a consistent sequence of confidence sets. Credible sets of minimal order (or diameter) equal $\{\theta_0\}$ with high $P_{\theta_{0,n}}$-probability.

Proof. Note that with uniform priors Π_n, $P_{\theta_{0,n}} \ll P_n^{\Pi}$ for all $n \ge 1$, so that P_n^{Π}-almost-surely defined credible sets D_n of credible level at least ϵ, also satisfy,

$$P_{\theta_{0,n}}\Big(\, \Pi(D_n(X^n)|X^n) \ge \epsilon \,\Big) = 1.$$

So if, in addition,

$$P_{\theta_{0,n}}\Big(\Pi(\{\theta_{0,n}\}|X^n) > 1-\epsilon \Big) \to 1,$$

then $\theta_{0,n} \in D_n(X^n)$ with high $P_{\theta_{0,n}}$-probability. Since all posterior mass is concentrated at $\theta_{0,n}$ with high probability, the $\{\theta_{0,n}\}$ form a sequence of unique credible sets of minimal order (or minimal diameter $k_n = 0$) with confidence levels greater than $\epsilon > 0$ for large enough n. □

When only *almost*-exact recovery is possible, the above strategy to obtain confidence sets, carries over for *enlargements of credible sets*. Recall the definition of the $V_{n,k}(\theta_n)$ in (5) (with $\theta_{0,n}$ replaced by θ_n). Given some fixed underlying $\theta_{0,n} \in \Theta_n$, we write $V_{n,k}$ for $V_{n,k}(\theta_{0,n})$. Making a certain choice for the upper bounds $k_n \geq 1$, we arrive at,

$$B_n(\theta_n) = \bigcup_{k=0}^{k_n} V_{n,k}(\theta_n), \tag{26}$$

for every $n \geq 1$ and $\theta_n \in \Theta_n$. Similar as for $V_{n,k}$ we write B_n for $B_n(\theta_{0,n})$. Given a subset D_n of Θ_n, the set $C_n \subset \Theta_n$ associated with D_n under $B_n(\theta_n)$ (see Definition B.2) then is the set of $\theta_n \in \Theta_n$ whose k-distance from some element of D_n is at most k_n,

$$C_n = \{\theta_n \in \Theta_n : \exists_{\eta_n \in D_n},\ k(\eta_n, \theta_n) \leq k_n\},$$

the k_n-enlargement of D_n. If we know that the sequences (p_n) and (q_n) satisfy requirement (23), posterior concentration occurs around $\{\theta_{0,n}\}$ in 'balls' of diameters $2k_n$ with growing n, and there exist credible sets D'_n of levels greater than $1/2$ and of diameters $2k_n$ centred on $\theta_{0,n}$. The credible sets D_n of *minimal diameters* of any level greater than $1/2$ must intersect D_n. Then the k_n-enlargements C_n of the D_n contain $\theta_{0,n}$.

Theorem 4.2. *Suppose that the posterior recovers communities almost-exactly with error rate (k_n),*

$$\Pi\Big(k(\theta_n, \theta_{0,n}) \leq k_n \ \Big|\ X^n \Big) \xrightarrow{P_{\theta_{0,n}}} 1.$$

Let (c_n) be bounded away from zero in the limit and let (D_n) denote a sequence of (P_n^{Π}-almost-sure) credible sets of levels c_n. Then the k_n-enlargements $C_n(X^n)$ of the $D_n(X^n)$ satisfy,

$$P_{\theta_{0,n}}\Big(\theta_{0,n} \in C_n(X^n) \Big) \to 1,$$

i.e., the k_n-enlargements (C_n) form a consistent sequence of confidence sets. If the sets D_n have minimal diameters, *then,*

$$\mathrm{diam}_n(D_n(X^n)) \leq 2k_n, \quad \mathrm{diam}_n(C_n(X^n)) \leq 4k_n,$$

with high $P_{\theta_{0,n}}$-probability.

Proof. As in the proof of Theorem 4.1, P_n^{Π}-almost-surely defined credible sets D_n of credible level at least c_n also satisfy,

$$P_{\theta_{0,n}}\Big(\Pi(D_n(X^n)|X^n) \geq c_n \Big) = 1.$$

Convergence of the posterior implies that with growing n, the balls $B_n(\theta_{0,n})$ of radii k_n centred on $\theta_{0,n}$ contain an arbitrarily large fraction of the total posterior mass, so assuming that n is large enough, $c_n > \epsilon > 0$ and $\Pi(B_n(\theta_{0,n})|X^n) > 1-\epsilon$ with high $P_{\theta_{0,n}}$-probability. Conclude that,

$$B_n(\theta_{0,n}) \cap D_n(X^n) \neq \varnothing,$$

with high $P_{\theta_{0,n}}$-probability, which amounts to asymptotic coverage of $\theta_{0,n}$ for the k_n-enlargement $C_n(X^n)$ of $D_n(X^n)$. Now fix $n \geq 1$. For every $\theta_n \in \Theta_n$ and every $x^n \in \mathscr{X}_n$, let $k_n(\theta_n, x^n)$ denote the minimal radius of balls B in Θ_n centred on θ_n of posterior mass $\Pi(B|x^n) \geq c_n$. Let $\hat{\theta}_n(x^n) \in \Theta_n$ be such that,

$$k_n(\hat{\theta}_n(x^n)) = \min\big\{ k_n(\theta_n, x^n) : \theta_n \in \Theta_n \big\},$$

i.e., the centre point of a *smallest* level-c_n credible ball in Θ_n. To conclude, note that $k_n(\hat{\theta}_n(X^n)) \leq k_n$ with high $P_{\theta_{0,n}}$-probability and if the $D_n(X^n)$ are of minimal diameters, then they are contained in $k_n(\hat{\theta}_n(X^n))$-balls centred on some $\hat{\theta}_n(X^n)$. □

4.2. Confidence sets directly from credible sets

To use Theorem 4.1 or Theorem 4.2, the statistician needs to know that the sequences (p_n) and (q_n) satisfy (12) or (23), basically to satisfy the testing condition (8). Particularly, condition (25) is *not* strong enough to apply Theorem 4.2. But even if that knowledge is not available and testing cannot serve as a condition, the use of credible sets as confidence sets remains valid, as long as credible levels grow to one fast enough. The following proposition also provides lower bounds for confidence levels of credible sets.

Proposition 4.3. *Let $\theta_{0,n}$ in Θ_n with uniform priors Π_n, $n \geq 1$, be given and define $b_n = |\Theta_n|^{-1} = (\frac{1}{2}\binom{2n}{n})^{-1}$. Let D_n be a sequence of credible sets, such that,*

$$\Pi(D_n(X^n)|X^n) \geq 1 - a_n,$$

for some sequence (a_n) with $a_n = o(b_n)$. Then,

$$P_{\theta_0,n}\big(\theta_0 \in D_n(X^n)\big) \geq 1 - b_n^{-1}a_n.$$

Proof. If $\theta_{0,n} \notin D_n(X^n)$ then $\Pi(\{\theta_{0,n}\}|X^n) \leq a_n$, P_n^{Π}-almost-surely. Then,

$$\begin{aligned} P_{\theta_0,n}\,\big(\theta_0 \in \Theta \setminus D_n(X^n)\,\big) &= P_n^{\Pi|\{\theta_0\}}\big(\theta_0 \in \Theta \setminus D_n(X^n)\big) \\ &= b_n^{-1}\int_{\{\theta_{0,n}\}} P_{\theta,n}\big(\theta_0 \in \Theta \setminus D_n(X^n)\big)\,d\Pi_n(\theta) \\ &= b_n^{-1}P_n^{\Pi}\big(1\{\theta_0 \in \Theta_n \setminus D_n(X^n)\}\,\Pi(\{\theta_{0,n}\}|X^n)\big) \leq b_n^{-1}a_n, \end{aligned}$$

by Bayes's Rule (A.3). □

Theorem B.3 leaves room for mitigation of the lower bound on credible levels if we are willing to use enlarged credible sets. There are two competing influences when enlarging: on the one hand, the prior masses $b_n = \Pi_n(B_n(\theta_{0,n}))$ become larger, relaxing the lower bounds for credible levels. On the other hand, enlargement leads to likelihood ratios with random fluctuations that take them further away from one (see Lemmas 4.4 and B.5), thus interfering with notions like contiguity and remote contiguity (see Appendix B and Kleijn (2021)). Whether Proposition 4.3 is useful and whether enlargement of credible sets helps, depends on the sequences (p_n) and (q_n): we consider the situation in which edge differences between within-community and between-community edge probabilities become less-and-less pronounced:

$$p_n - q_n = o\big(n^{-1}\big), \tag{27}$$

while satisfying also the condition that,

$$p_n^{1/2}(1-p_n)^{1/2} + q_n^{1/2}(1-q_n)^{1/2} = o\big(n|p_n - q_n|\big). \tag{28}$$

In this regime either $p_n, q_n \to 0$ or $p_n, q_n \to 1$, signifying sparsity of either presence or absence of edges respectively. If $p_n, q_n \to 0$, (28) amounts to,

$$n(p_n^{1/2} - q_n^{1/2}) \to \infty, \tag{29}$$

so differences between p_n and q_n may not converge to zero too fast (essentially in order to maintain sufficient distinction from the Erdős-Rényi graph Janson, 2010; Banerjee, 2018). For the following lemma we define,

$$\rho_n = \min\left\{\left(\frac{1-p_n}{p_n}\,\frac{q_n}{1-q_n}\right), \left(\frac{p_n}{1-p_n}\,\frac{1-q_n}{q_n}\right)\right\} = e^{-|\lambda_n|},$$

where $\lambda_n := \log(1-p_n) - \log(p_n) + \log(q_n) - \log(1-q_n)$, and,

$$\alpha_n = \int 2k(\theta_{0,n},\theta_n)(n - k(\theta_{0,n},\theta_n))\,d\Pi_n(\theta_n|B_n) = \frac{1}{|B_n|}\sum_{k=0}^{k_n}\binom{n}{k}^2 2k(n-k),$$

with the following rate for remote contiguity (see Definition B.4):

$$d_n = \rho_n^{C\alpha_n|p_n - q_n|}, \tag{30}$$

for some $C > 1$.

For the following lemma, let (p_n) and (q_n) be given (with resulting (ρ_n)), and let $C >$, (k_n) and $\theta_{0,n} \in \Theta_n$, for all $n \geq 1$ be given (with resulting (α_n) and (d_n)).

Lemma 4.4. *Assume that (28) holds. Then,*

$$P_{\theta_0,n} \lhd d_n^{-1}P_n^{\Pi|B_n},$$

with $B_n = B_n(\theta_{0,n})$ like in (26).

Proof. Let (k_n) and $\theta_{0,n} \in \Theta_n$ be given. We denote $P_n = P_n^{\Pi|B}$, $Q_n = P_{\theta_0,n}$ and apply Jensen's inequality to obtain,

$$\begin{aligned} \frac{dP_n}{dQ_n}(X^n) &= \frac{1}{|B_n|}\sum_{\theta_n \in B_n}\left(\frac{1-p_n}{p_n}\,\frac{q_n}{1-q_n}\right)^{S_n(\theta_n)-T_n(\theta_n)} \\ &\geq \exp\Big(\frac{\lambda_n}{|B_n|}\sum_{\theta_n \in B_n}\big(S_n(\theta_n) - T_n(\theta_n)\big)\Big), \end{aligned}$$

where $(S_n(\theta_n), T_n(\theta_n))$ is distributed as in (11). By invariance of the sum under permutations of the vertices, we re-sum as follows for any $k \geq 1$,

$$\frac{1}{|V_{n,k}|}\sum_{\theta_n\in V_{n,k}} S_n(\theta_n) = \frac{2k(n-k)}{n(n-1)}S_n, \quad \frac{1}{|V_{n,k}|}\sum_{\theta_n\in V_{n,k}} T_n(\theta_n) = \frac{2k(n-k)}{n^2}T_n,$$

where, with the notation $Z_n = Z(\theta'_{0,n}) \subset \{1,\ldots,2n\}$, for a certain representation $\theta'_{0,n}$ of $\theta_{0,n}$, for the zero elements of $\theta'_{0,n}$,

$$S_n = \sum_{i,j\in Z_n} X_{ij} + \sum_{i,j\in Z_n^c} X_{ij} \sim \mathrm{Bin}(n(n-1), p_n),$$
$$T_n = \sum_{i\in Z_n, j\in Z^c} X_{ij} + \sum_{i\in Z_n^c, j\in Z} X_{ij} \sim \mathrm{Bin}(n^2, q_n),$$

which gives us the lower bound,

$$\frac{dP_n}{dQ_n}(X^n) \geq \rho_n^{\sum_{k=0}^{k_n} 2k(n-k)\frac{|V_{n,k}|}{|B_n|}|\bar{S}_n-\bar{T}_n|} = \rho_n^{\alpha_n|\bar{S}_n-\bar{T}_n|},$$

where $\bar{S}_n = S_n/(n(n-1))$ and $\bar{T}_n = T_n/n^2$. By the central limit theorem,

$$\left(\frac{n(\bar{S}_n-p_n)}{p_n^{1/2}(1-p_n)^{1/2}}, \frac{n(\bar{T}_n-q_n)}{q_n^{1/2}(1-q_n)^{1/2}}\right) \xrightarrow{Q_n\text{-w.}} N(0,1)\times N(0,1),$$

which implies that for every $\epsilon > 0$ there exists an $M > 0$ such that,

$$\sup_{n\geq 1} Q_n\left(\frac{n(\bar{S}_n-p_n)}{p_n^{1/2}(1-p_n)^{1/2}} \vee \frac{n(\bar{T}_n-q_n)}{q_n^{1/2}(1-q_n)^{1/2}} > M\right) < \epsilon.$$

Conclude that,

$$\sup_{n\geq 1} Q_n\left(\left(\frac{dP_n}{dQ_n}\right)^{-1} \leq \rho_n^{-\alpha_n\left(\frac{M}{n}(p_n^{1/2}(1-p_n)^{1/2}+q_n^{1/2}(1-q_n)^{1/2})+|p_n-q_n|\right)}\right) \geq 1-\epsilon.$$

Note that the term in the exponent proportional to M is dominated by $|p_n - q_n|$ by (28). Hence for every $C > 1$ and every $\epsilon > 0$,

$$Q_n\left(\left(\frac{dP_n}{dQ_n}(X^n)\right)^{-1} \leq \rho_n^{-C\alpha_n|p_n-q_n|}\right) \geq 1-\epsilon,$$

for large enough n. Using the remark following Lemma B.5, we see that $P_{\theta_{0,n}} \lhd d_n^{-1}P_n^{\Pi|B_n}$, with d_n as in (30). □

This argument amounts to a proof for the following theorem (immediate from Theorem B.3).

Theorem 4.5. *Let (k_n) be given and assume that (p_n) and (q_n) satisfy (27) and (28) (or $p_n, q_n \to 0$ and (29)). Let $\theta_{0,n}$ in Θ_n with uniform priors Π_n be given and let $D_n(X^n)$ be a sequence of credible sets of credible levels $1-a_n$, for some sequence (a_n) such that $b_n^{-1}a_n = o(d_n)$. Then the sets $C_n(X^n)$, associated with $D_n(X^n)$ under B_n as in (26) satisfy,*

$$P_{\theta_{0,n}}\big(\theta_0 \in C_n(X^n)\big) \to 1,$$

i.e., the $C_n(X^n)$ are asymptotic confidence sets.

Consider the possible choices for (a_n) if we assume $k_n = \beta\, n$ for some fixed $\beta \in (0, \frac{1}{2})$ (as in the proof of Corollary 3.5). First of all, Stirling's approximation gives rise to the following approximate lower bound on the factor between prior mass and prior mass without enlargement:

$$\frac{\Pi_n(B_n)}{\Pi_n(\{\theta_{0,n}\})} = \sum_{k=0}^{k_n}\binom{n}{k}^2 \geq \binom{n}{k_n}^2 \geq \frac{1}{2\pi n}\frac{1}{\beta(1-\beta)}f(\beta)^n,$$

where $f : (0, \frac{1}{2}) \to (1, 4)$ is given by,

$$f(\beta) = (1-\beta)^{-2(1-\beta)}\beta^{-2\beta}.$$

Approximating $\alpha_n \approx 2k_n(n-k_n)$ for large n and using (27), we also have,

$$d_n = \rho_n^{C\alpha_n|p_n-q_n|} \approx \rho_n^{2Cn^2\beta(1-\beta)|p_n-q_n|} = e^{-|\lambda_n|o(n)}.$$

So if we assume that $\lambda_n = O(1)$, d_n is sub-exponential and does not play a role for the improvement factor.

Conclude as follows: (let $a_n = o(|\Theta_n|^{-1}) \approx o(4^{-n})$ denote the rates appropriate in Proposition 4.3 and assume $\lambda_n = O(1)$) if we have credible sets $D_n(X^n)$ of credible levels $1 - a_n f(\beta)^{n(1+o(1))}$, then the sequence of enlarged confidence sets $(C_n(X^n))$, associated with $D_n(X^n)$ through B_n with $k_n = \beta n$, covers the true value of the community assignment parameter with high probability. Credible levels that had to be of order $1 - a_n \approx 1 - o(4^{-n})$ previously, can be of approximate order $1 - o(c^{-n})$ for any $1 < c < 4$ by enlargement by B_n if conditions (27) and (28) hold; the closer $0 < \beta < \frac{1}{2}$ is to $\frac{1}{2}$, the closer c is to 1.

5. Conclusions and discussion

In this paper we consider application of Bayesian posteriors for frequentist asymptotic inference on the community structure of sparse planted bi-section graphs. More specifically, we prove that the posterior recovers the true community assignment (exactly, respectively almost-exactly) in the sparsest possible (Chernoff–Hellinger, respectively Kesten-Stigum) cases.

We also use the posterior concentration results to draw conclusions regarding the role of (enlarged) Bayesian credible sets as frequentist asymptotic confidence sets. For this purpose, it is important that the posterior concentration results are sharp; otherwise the credible sets we choose are too large and the required confidence levels are too high, leading to asymptotic confidence sets that are too conservative.

The analysis we give is limited in several respects. First of all, although realistic regarding expected degrees in large graphs, the planted bi-section model is a highly stylized random graph model; more flexible is the family of stochastic block models, which leaves room for more than two communities, and classes of unequal sizes (and generalizations thereof). On uncertainty quantification in stochastic block models, the literature is very limited: in van Waaij and Kleijn (2020), the present analysis is extended to an unknown number of communities of order $O(\sqrt{n/\log(n)})$, of unknown sizes bounded above and below proportional to the graph size n.

Practical implementation of what we propose is not straightforward: with a graph X^n of fixed size n, a bound like (9) in combination with Lemma 2.5 (both of which hold for finite n) permits calculation of a lower bound for k_n. With the corresponding enlargement radius, confidence sets can then be constructed as explicit radius-k_n-enlargements of credible sets from the posterior. This finite-n programme is followed in Kleijn and van Waaij (2021).

To conclude we note that Bayesian methods may have lost some of their popularity of late, because the computational burden of sampling a posterior distribution is deemed relatively high. In the planted bi-section model, for example, other more efficient methods for the recovery of the community structure in the planted bi-section model exist. Based on the preceding, however, we argue that if *uncertainty quantification* for the community structure is the goal, the relatively high computational cost of simulating a posterior is justifiable. Since (limiting) sampling distributions of other estimators are prohibitively hard to obtain or analyse, constructing asymptotic confidence sets for community structure in other ways may prove to be very hard or even impossible.

Acknowledgements

The authors thank E. Mossel and J. Neeman for a helpful discussion on necessary conditions for exact recovery. BK thanks P. Bickel for his encouragement to pursue the confidence-sets-from-credible-sets question.

Appendix A. Definitions and conventions

We assume given for every $n \geq 1$, a random graph X^n taking values in the (finite) space $\mathscr{X}_n$ of all undirected graphs with n vertices and no self-loops. We denote the powerset of $\mathscr{X}_n$ by $\mathscr{B}_n$ and regard it as the domain for probability distributions $P_n : \mathscr{B}_n \to [0, 1]$ a model $\mathscr{P}_n$ parametrized by $\Theta_n \to \mathscr{P}_n : \theta \mapsto P_{\theta,n}$ with finite parameter spaces Θ_n (with powerset $\mathscr{G}_n$) and uniform priors Π_n on Θ_n. As frequentists, we assume that there exists a 'true, underlying distribution for the data' $P_{0,n}$; in this case, that means that for every $n \geq 1$, there exists a $\theta_{0,n} \in \Theta_n$ and corresponding $P_{\theta_0,n}$ from which the nth graph X^n is drawn.

Definition A.1. Given $n \geq 1$ and a prior probability measure Π_n on Θ_n, define the *nth prior predictive distribution* as:

$$P_n^{\Pi}(A) = \int_{\Theta} P_{\theta,n}(A)\, d\Pi_n(\theta), \tag{A.1}$$

for all $A \in \mathscr{B}_n$. For any $B_n \in \mathscr{G}_n$ with $\Pi_n(B_n) > 0$, define also the *nth local prior predictive distribution*,

$$P_n^{\Pi|B}(A) = \frac{1}{\Pi_n(B_n)} \int_{B_n} P_{\theta,n}(A)\, d\Pi_n(\theta), \tag{A.2}$$

as the predictive distribution on $\mathscr{X}_n$ that results from the prior Π_n when conditioned on B_n.

The prior predictive distribution P_n^{Π} is the marginal distribution for X^n in the Bayesian perspective that considers parameter and sample jointly $(\theta, X^n) \in \Theta \times \mathscr{X}_n$ as the random quantity of interest.

Definition A.2. Given $n \geq 1$, a *(version of) the posterior* is any set-function $\mathscr{G}_n \times \mathscr{X}_n \to [0,1] : (A, x^n) \mapsto \Pi(A\,|X^n = x^n)$ such that,

1. for $B \in \mathscr{G}_n$, the map $x^n \mapsto \Pi(B|X^n = x^n)$ is $\mathscr{B}_n$-measurable,
2. for all $A \in \mathscr{B}_n$ and $V \in \mathscr{G}_n$,

$$\int_A \Pi(V|X^n)\, dP_n^{\Pi} = \int_V P_{\theta,n}(A)\, d\Pi_n(\theta). \tag{A.3}$$

Bayes's Rule is expressed through equality (A.3) and is sometimes referred to as a 'disintegration' (of the joint distribution of (θ, X^n)).

Because we take the perspective of a frequentist using Bayesian methods, we are obliged to demonstrate that Bayesian definitions continue to make sense under the assumption that the data X^n is distributed according to a true, underlying $P_{0,n}$: Bayesian concepts above that have been defined through conditioning, are almost-sure with respect to the relevant marginal. In the case of X^n, the relevant marginal is the prior predictive distribution. Accordingly, we have to assume that $P_{\theta_0,n} \ll P_n^{\Pi}$, for all $n \geq 1$.

The following lemma is used in two places in the text.

Lemma A.3. *For all positive integers r and real numbers $x > -r$, $(1 + x/r)^r \leq e^x$.*

Proof. Let for $x > -r$, $f(x) = r\log(1+x/r)$ and $g(x) = x$. Then $f'(x) = (1+x/r)^{-1}$ and $g'(x) = 1$. Then $f'(x) \leq g'(x)$, when $x \geq 0$, $f'(x) > g'(x)$ when $-n < x < 0$ and $f(0) = g(0)$. It follows that $f(x) \leq g(x)$ for all $x > -r$. As $y \to e^y$ is increasing for all real y, we find $x > -n$, $(1 + x/r)^r = e^{f(x)} \leq e^{g(x)} = e^x$. $\square$

Notation and conventions

Asymptotic statements that end in "... with high probability" indicate that said statements are true with probabilities that grow to one. For given probability measures P, Q on a measurable space $(\Omega, \mathscr{F})$, we define the Radon–Nikodym derivative $dP/dQ : \Omega \to [0,\infty)$, P-almost-surely, referring *only* to the Q-dominated component of P, following Le Cam (1986). We also *define* $(dP/dQ)^{-1} : \Omega \to (0,\infty] : \omega \mapsto 1/(dP/dQ(\omega))$, Q-almost-surely. Given random variables $Z_n \sim P_n$, weak convergence to a random variable Z is denoted by $Z_n \xrightarrow{P_n\text{-w.}} Z$, convergence in probability by $Z_n \xrightarrow{P_n} Z$ and almost-sure convergence (with coupling P^∞) by $Z_n \xrightarrow{P^\infty\text{-a.s.}} Z$. The integral of a real-valued, integrable random variable X with respect to a probability measure P is denoted PX, while integrals over the model with respect to priors and posteriors are always written out in Leibniz's or sum notation. The cardinality of a set B is denoted $|B|$.

Appendix B. Remote contiguity and confidence sets

Bayesian asymptotics has seen a great deal of development over recent decades, but the essence of the theory remains that of Schwartz's theorem: a balance between testing power and a minimum of prior mass 'locally', leads to a controlled limit for the posterior distribution with a frequentist interpretation. It has also become clear that the same notion of 'locality' allows conversion of sequences of credible sets to asymptotic confidence sets and that is the purpose of this paper as well. 'Locality' in the above sense is defined through a weakened form of contiguity called remote contiguity (Kleijn, 2021). In this appendix, we summarize these points, to support the proofs of Lemma 4.4 and Theorem 4.5.

Definition B.1. Let $(\Theta_n, \mathscr{G}_n)$ with priors Π_n be given, denote the sequence of posteriors by $\Pi(\cdot|\cdot) : \mathscr{G}_n \times \mathscr{X}_n \to [0,1]$. Let $\mathscr{D}_n$ denote a collection of measurable subsets of Θ_n. A *sequence of credible sets* (D_n) of *credible levels* $1 - a_n$ (where $0 \leq a_n \leq 1$, $a_n \downarrow 0$) is a sequence of set-valued maps $D_n : \mathscr{X}_n \to \mathscr{D}_n$ such that $\Pi(\Theta_n \setminus D_n(x^n)|X^n = x^n) \leq a_n$.

Note that the posterior is defined P_n^{Π}-almost-surely with respect to its dependence on the data X^n, and consequently, so is any credible set that is derived from it.

Definition B.2. Let D be a (credible) set in Θ and let $B = \{B(\theta) : \theta \in \Theta\}$ denote a collection of model subsets such that $\theta \in B(\theta)$ for all $\theta \in \Theta$. A model subset C is said to be (a confidence set) associated with D under B, if for all $\theta \in \Theta \setminus C$, $B(\theta) \cap D = \varnothing$.

The relationship between a credible set D and the model subset C associated with D under B is illustrated in Fig. B.3 (reproduced from Kleijn, 2021, Figure 1) and detailed in the following theorem. (The notation $P_n \lhd d_n^{-1} Q_n$ is explained below, see Definition B.4.)

Theorem B.3. *Let $\theta_{0,n} \in \Theta_n$ $(n \geq 1)$ and $0 \leq a_n \leq 1$, $b_n > 0$ such that $a_n = o(b_n)$ be given. Choose priors Π_n and let D_n denote level-$(1 - a_n)$ credible sets in Θ_n. Furthermore, for all $\theta \in \Theta$, let $B_n = \{B_n(\theta_n) \in \mathscr{G}_n : \theta_n \in \Theta_n\}$ and b_n denote sequences such that,*

(i.) prior mass is lower bounded, $\Pi_n(B_n(\theta_0)) \geq b_n$,

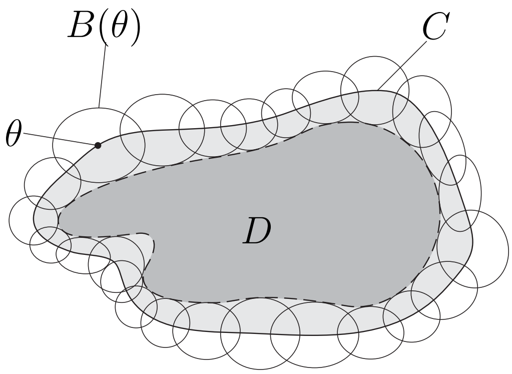

Fig. B.3. The relation between a credible set D and its associated confidence set C under B in Venn diagrams: the extra points θ in the associated confidence set C not included in the credible set D are characterized by non-empty intersection $B(\theta) \cap D \neq \varnothing$.
Source: Reproduced from Kleijn (2021, Figure 1).

(ii.) and for some $d_n \downarrow 0$ such that $b_n^{-1} a_n = o(d_n)$, $P_{\theta_0,n} \lhd d_n^{-1}\, P_n^{\Pi|B(\theta_0)}$.

Then any confidence sets C_n associated with the credible sets D_n under B_n are asymptotically consistent, i.e., for all $\theta_0 \in \Theta$,

$$P_{\theta_0,n}\big(\theta_0 \in C_n(X^n)\big) \to 1. \tag{B.1}$$

In most of Section 4, the sets B_n are simply,

$$B_n(\theta_n) = \{\theta_n\},$$

for every $n \geq 1$ and every $\theta_n \in \Theta_n$, so that the confidence sets C_n associated with any credible sets $D_n \subset \Theta_n$ under B_n are simply *equal to* D_n. In that case, $P_{\theta_0,n} \lhd c_n^{-1}\, P_n^{\Pi|B(\theta_0)}$ for any rate (c_n), $c_n \downarrow 0$, so all sequences $a_n = o(b_n)$ are permitted. Since the prior mass in $B_n(\theta_{0,n})$ is fixed, Theorem B.3 says that, if we have a sequence of credible sets $D_n(X^n) \subset \Theta_n$ of high enough credible levels $1 - a_n$, then these $D_n(X^n)$ are also asymptotically consistent confidence sets (see Proposition 4.3).

With an eye on enlarged credible sets, we note that condition *(ii.)* of Theorem B.3 says that the sequence $(P_{\theta_0,n})$ is required to be *remotely contiguous* with respect to $P_n^{\Pi|B(\theta_0)}$ at rate $b_n a_n^{-1}$.

Definition B.4. Given the spaces $\mathscr{X}_n$, $n \geq 1$ with two sequences (P_n) and (Q_n) of probability measures and a sequence $\rho_n \downarrow 0$, we say that Q_n is ρ_n-remotely contiguous with respect to P_n, notation $Q_n \lhd \rho_n^{-1} P_n$, if,

$$P_n\phi_n(X^n) = o(\rho_n) \quad \Rightarrow \quad Q_n\phi_n(X^n) = o(1), \tag{B.2}$$

for every sequence of $\mathscr{B}_n$-measurable $\phi_n : \mathscr{X}_n \to [0,1]$.

According to section 3 in Kleijn (2021), weak relative compactness of a sequence of re-scaled (inverse) likelihood ratios is sufficient for remote contiguity.

Lemma B.5. *Given (P_n), (Q_n), $d_n \downarrow 0$, (Q_n) is d_n-remotely contiguous with respect to (P_n) if, under Q_n, every subsequence of $(d_n(dP_n/dQ_n)^{-1})$ has a weakly convergent subsequence.*

According to Prokhorov's theorem, the condition of Lemma B.5 is equivalent to uniform tightness: for every $\epsilon > 0$ there exists an $M > 0$ such that,

$$\sup_{n\geq 1} P\Big(d_n\Big(\frac{dP_n}{dQ_n}\Big)^{-1}(X^n) > M\Big) < \epsilon.$$

References

Abbe, E., 2018. Community detection and stochastic block models: Recent developments. J. Mach. Learn. Res. 18 (177), 1–86, URL: http://jmlr.org/papers/v18/16-480.html.

Abbe, E., Bandeira, A., Hall, G., 2016. Exact recovery in the stochastic block model. IEEE: Trans. Inf. Theory 62 (1), URL: http://arxiv.org/abs/1405.3267.

Amini, A., Chen, A., Bickel, P., Levina, E., 2013. Pseudo-likelihood methods for community detection in large sparse networks. Ann. Statist. 41 (4), 2097–2122. http://dx.doi.org/10.1214/13-AOS1138.

Banerjee, D., 2018. Contiguity and non-reconstruction results for planted partition models: the dense case. Electron. J. Probab. 23, 28. http://dx.doi.org/10.1214/17-EJP128.

Bickel, P., Chen, A., 2009. A nonparametric view of network models and Newman-Girvan and other modularities. Proc. Natl. Acad. Sci. 106 (50), 21068–21073. http://dx.doi.org/10.1073/pnas.0907096106.

Bickel, P., Sarkar, P., 2016. Hypothesis testing for automated community detection in networks. J. R. Stat. Soc. Ser. B 78 (1), 253–273, URL: https://ideas.repec.org/a/bla/jorssb/v78y2016i1p253-273.html.
Bollobás, B., Janson, S., Riordan, O., 2007. The phase transition in inhomogeneous random graphs. Random Struct. Algorithms 31 (1), 3–122. http://dx.doi.org/10.1002/rsa.20168.
Choi, D., Wolfe, P., Airoldi, E., 2012. Stochastic blockmodels with a growing number of classes. Biometrika 99 (2), 273–284. http://dx.doi.org/10.1093/biomet/asr053.
Decelle, A., Krzakala, F., Moore, C., Zdeborová, L., 2011a. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. Phys. Rev. E 84, 066106. http://dx.doi.org/10.1103/PhysRevE.84.066106.
Decelle, A., Krzakala, F., Moore, C., Zdeborová, L., 2011b. Inference and phase transitions in the detection of modules in sparse networks. Phys. Rev. Lett. 107, 065701. http://dx.doi.org/10.1103/PhysRevLett.107.065701.
Dyer, M., Frieze, A., 1989. The solution of some random NP-hard problems in polynomial expected time. J. Algorithms 10 (4), 451–489. http://dx.doi.org/10.1016/0196-6774(89)90001-1.
Erdős, P., Rényi, A., 1959. On random graphs I. Publ. Math..
Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486 (3), 75–174. http://dx.doi.org/10.1016/j.physrep.2009.11.002.
Gao, C., Ma, Z., Zhang, A., Zhou, H., 2017. Achieving optimal misclassification proportion in stochastic block models. J. Mach. Learn. Res. 18 (60), 1–45, URL: http://jmlr.org/papers/v18/16-245.html.
Geng, J., Bhattacharya, A., Pati, D., 2019. Probabilistic community detection with unknown number of communities. J. Amer. Statist. Assoc. 114 (526), 893–905, URL: https://EconPapers.repec.org/RePEc:taf:jnlasa:v:114:y:2019:i:526:p:893-905.
Girvan, M., Newman, M., 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA 99 (12), 7821–7826. http://dx.doi.org/10.1073/pnas.122653799.
Guédon, O., Vershynin, R., 2016. Community detection in sparse networks via Grothendieck's inequality. Probab. Theory Related Fields 165 (3), 1025–1049. http://dx.doi.org/10.1007/s00440-015-0659-z.
Hajek, B., Wu, Y., Xu, J., 2016. Achieving exact cluster recovery threshold via semidefinite programming. IEEE Trans. Inf. Theor. 62 (5), 2788–2797. http://dx.doi.org/10.1109/TIT.2016.2546280.
Holland, P., Laskey, K., Leinhardt, S., 1983. Stochastic blockmodels: First steps. Social Networks 5 (2), 109–137. http://dx.doi.org/10.1016/0378-8733(83)90021-7.
Janson, S., 2010. Asymptotic equivalence and contiguity of some random graphs. Random Struct. Algorithms 36 (1), 26–45. http://dx.doi.org/10.1002/rsa.20297.
Jiang, S., Tokdar, S., 2021. Consistent Bayesian community detection. arXiv:2101.06531.
Kleijn, B., 2021. Frequentist validity of Bayesian limits. Ann. Statist. 49 (1), 182–202. http://dx.doi.org/10.1214/20-AOS1952.
Kleijn, B.J.K., van Waaij, J., 2021. Confidence sets in a sparse stochastic block model with two communities of unknown sizes. http://dx.doi.org/10.48550/ARXIV.2108.07078.
Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., Zhang, P., 2013. Spectral redemption in clustering sparse networks. Proc. Natl. Acad. Sci. 110 (52), 20935–20940. http://dx.doi.org/10.1073/pnas.1312486110.
Le Cam, L., 1986. Asymptotic Methods in Statistical Decision Theory. Springer-Verlag, New York, http://dx.doi.org/10.1007/978-1-4612-4946-7.
Lei, J., 2016. A goodness-of-fit test for stochastic block models. Ann. Statist. 44 (1), 401–424. http://dx.doi.org/10.1214/15-AOS1370.
Massoulié, L., 2014. Community detection thresholds and the weak Ramanujan property. In: STOC 2014: 46th Annual Symposium on the Theory of Computing. New York, United States, pp. 1–10, URL: https://hal.archives-ouvertes.fr/hal-00969235.
McDaid, A., Murphy, T., Friel, N., Hurley, N., 2013. Improved Bayesian inference for the stochastic block model with application to large networks. Comput. Statist. Data Anal. 60 (C), 12–31. http://dx.doi.org/10.1016/j.csda.2012.10.02.
Mossel, E., Neeman, J., Sly, A., 2015. Reconstruction and estimation in the planted partition model. Probab. Theory Related Fields 162 (3), 431–461. http://dx.doi.org/10.1007/s00440-014-0576-6.
Mossel, E., Neeman, J., Sly, A., 2016. Consistency thresholds for the planted bisection model. Electron. J. Probab. 21, 24. http://dx.doi.org/10.1214/16-EJP4185.
Nowicki, K., Snijders, T., 2001. Estimation and prediction for stochastic blockstructures. J. Amer. Statist. Assoc. 96 (455), 1077–1087. http://dx.doi.org/10.1198/016214501753208735.
Suwan, S., Lee, D., Tang, R., Sussman, D., Tang, M., Priebe, C., 2016. Empirical Bayes estimation for the stochastic blockmodel. Electron. J. Stat. 10 (1), 761–782. http://dx.doi.org/10.1214/16-EJS1115.
van Waaij, J., Kleijn, B.J.K., 2020. Uncertainty quantification in the stochastic block model with an unknown number of classes. http://dx.doi.org/10.48550/ARXIV.2005.01362.
Yuan, M., Feng, Y., Shang, Z., 2022. A likelihood-ratio type test for stochastic block models with bounded degrees. J. Stat. Plan. Inference 219, 98–119. http://dx.doi.org/10.1016/j.jspi.2021.12.005.
Zhang, A., Zhou, H., 2016. Minimax rates of community detection in stochastic block models. Ann. Statist. 44 (5), 2252–2280. http://dx.doi.org/10.1214/15-AOS1428.