

Inference in the Stochastic Block Model with a Markovian assignment of the communities.

Quentin Duchemin*

April 2020

Abstract

We tackle the community detection problem in the Stochastic Block Model (SBM) when the communities of the nodes of the graph are assigned with a Markovian dynamic. To recover the partition of the nodes, we adapt the relaxed K -means SDP program presented in [11]. We identify the relevant signal-to-noise ratio (SNR) in our framework and we prove that the misclassification error decays exponentially fast with respect to this SNR. We provide infinity norm consistent estimation of the parameters of our model and we discuss our results through the prism of classical degree regimes of the SBMs' literature.

MSC 2010 subject classifications: Primary 68Q32; secondary 68R10, 90C35.

Keywords: Random graphs, Stochastic Block Models, Markov chains, Semidefinite relaxation.

1 Introduction

Large random graphs have been very popular in the last decade since they are powerful tools to model complex phenomena like interactions on social networks or the spread of a disease. In practical cases, detecting communities of well connected nodes in a graph is a major issue, motivating the study of the Stochastic Block Model (SBM). In this model, each node belongs to a particular community and edges are sampled independently according to a probability depending of the communities of the nodes. Aiming at progressively bridging the gap between models and reality, time evolving random graphs have been recently introduced. In [20], a Stochastic Block Temporal Model is considered where the temporal evolution is modeled through a discrete hidden Markov chain on the nodes membership and where the connection probabilities also evolve through time. In [22], connection probabilities between nodes are functions of time, considering a maximum number of nodes that can switch their communities between two consecutive time steps. Following the work of [17], [19] study the Degree Corrected Stochastic Block Model where the degree of the nodes can vary within the same community. They show that for the relatively sparse case (i.e. when the maximum expected node degree is of order $\log(n)$ or higher), the proportion of misclassified nodes tends to 0 with a probability that goes to 1 when the number of nodes n increases using spectral clustering. This result inspired the recent paper [18] which considers a Dynamic Stochastic Block Model where the communities can change with time. They provide direct link between the density of the graph and its smoothness (which measures how much the graph change with time): the smoother is the graph, the sparser it can be guaranteeing still consistent recovery of communities. Several other dynamic variants of the SBM have been proposed so far like in [25] where the presence of an edge at the time step $t + 1$ directly depends on its presence or absence at time t .

Overview of resolution tools for the static SBM A large span of methods have been developed to solve clustering problems on graphs. The survey [1] gathers the state of the art methods to solve the community detection problem in the SBM. Abbe proposes a well-structured paper treating a large number of different settings for the SBM and reveals phase transition phenomena. Different recovery requirements have been studied in the SBM. Exact recovery defines the ability to recover the true partition of the nodes as the size of graph tends to $+\infty$ while partial recovery aims at asymptotically recovering correctly a fixed and a non trivial fraction of the partition of the nodes. If the survey focuses mainly

*LAMA, Univ Gustave Eiffel, Univ Paris Est Créteil, CNRS, F-77447 Marne-la-Vallée, France. quentin.duchemin@univ-eiffel.fr

on the symmetric (i.e. with balanced communities) SBM with two communities, results concerning the general SBM are also presented. Depending on the recovery requirement and the framework, different tools have been used to tackle the community detection problem. In [2], an algorithm based on belief propagation solves weak recovery in the general SBM with $K \geq 2$ communities and a signal-to-noise ratio (SNR) larger than 1. In [7] a spectral method is used to solve almost exact recovery in the SBM. If neural networks [24], Bayesian approaches [26] or Maximum Likelihood estimation [4] have also been proposed to address the community detection problem, another powerful and popular method is Semi-Definite Programming (SDP).

SDPs in the SBM literature If SDP methods can be used to solve weak or exact recovery, it is now well-known that they are usually not the correct tool to reach the optimal threshold [15], [23]. Despite this lack of optimality, SDP methods have interesting robustness features [23], [9]. Nevertheless, SDP methods studied so far are suffering from two main drawbacks as pointed out by [11]. i) Most of them need the communities to be balanced or to know their sizes ii) Most of them are based on convex relaxations of min-cut optimization problems [6], [23], [13], [9] and thus can only be used for the assortative SBM¹. Giraud and Verzelen propose in [11] a SDP method that overcomes the two obstacles previously highlighted. They solve a relaxed version of K -means to get partial recovery bound with a misclassification error that decays exponentially fast with the SNR. Their result, inspired from [9], is improving the partial recovery bounds previously obtained with the Groethendieck inequality in [12] and [8].

Contributions Contrary to the dynamic SBMs already presented in the literature, we introduce a new model where the size of the graph grows with time. If previous works mainly consider a fixed number of nodes with an evolving graph where at each time step communities or connection probabilities can change, our point of view is different. In our model, a new node is added to the graph at each time point. We assign a community using a Markovian transition starting from the community of the last node inserted. The newcomer is then connected to all the other nodes of the graph with some probability. Figure 1 gives a graphical representation of our model. In this work, we use the algorithm from [11] to tackle community detection in graphs sampled from our new SBM model. We prove consistent estimation of the parameters of the model using the infinity norm. We comment our results based on the classical degree regimes in the SBMs' literature.

Connection with the litterature If our model shares similarities with Hidden Markov Model (HMM), it turns out to be richer because the emission probabilities are not independent. This is due to the fact that each new node will be connected to all the other nodes of the graph with some probabilities depending on the communities. A classical approach to tackle estimation of the hidden variables in HMMs is the Expectation-Maximization (EM) algorithm. However this method suffers from a lack of theoretical guarantees on the convergence state. In particular the limit highly depends on the initialization of the algorithm. In [14] a spectral algorithm is used to learn HMMs. They provide a guarantee to recover the joint law of the observations with their polynomial time algorithm. Their result holds under the assumption that the transition matrix of the hidden chain is non singular, and that the (finitely valued) emission distributions are linearly independent. Under those conditions, they ensure the identifiability of their model. [10] provides, under the same two conditions, the general identifiability result for finite state space HMMs. In Section 2, we show that our model is identifiable when the node average degree is of order $\log(n)$ or higher.

Outline In Section 2, we start by defining the SBM and its modified version with a Markovian assignment of the communities. In Section 3, we describe the algorithm from [11] leading to a partial recovery bound for a deterministic assignment of the communities. In Section 4, we provide the explicit way to estimate the parameters of our model. In Section 5, we expose our results. After presenting some numerical experiments in Section 6, the proofs are provided in Section 7.

Notations Let $p \in \mathbb{N}^*$.

- $[p] := \{1, \dots, p\}$.
- $\forall A, B \in \mathbb{R}^{p \times p}, \quad \langle A, B \rangle = \text{Tr}(A^\top B)$.

¹We recall that the assortative SBM refers to the case where within group probabilities of connection are larger than between group probabilities of connection.

-
- $\forall a, b \in \mathbb{R}, \quad a \wedge b := \min(a, b)$ and $a \vee b := \max(a, b)$.
 - $\forall A \in \mathbb{R}^{p \times p}, \quad |A|_1 := \sum_{1 \leq i, j \leq p} |A_{i,j}|$.
 - $\forall A \in \mathbb{R}^{p \times p}, \quad \forall i \in [p], A_{:,i}$ denotes the i -th column of the matrix A .
 - $\forall u \in \mathbb{R}^p, \quad \|u\| := \sqrt{\sum_{i=1}^p u_i^2}$.
 - For a varying parameter α_p , the notation $\alpha_p \sim f(p)$ indicates that, as $p \rightarrow \infty$, the quantity $\alpha_p/f(p)$ tends to a non-zero constant.
 - For a varying parameter α_p , the notation $\alpha_p = \Omega(f(p))$ indicates that, as $p \rightarrow \infty$, the quantity $|f(p)/\alpha_p|$ stays bounded.

2 Model

An undirected graph G is defined by a set of nodes V and a set of edges $E \subset V \times V$. For an undirected graph with n nodes, we define the adjacency matrix of this graph $X \in \{0, 1\}^{n \times n}$ such that for all $i, j \in [n]$,

$$X_{i,j} = \begin{cases} 1 & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise.} \end{cases}$$

Stochastic Block Model Let us consider $K \geq 2$ communities and a set of n nodes $V = [n]$. Each node $i \in [n]$ belongs to one community $k \in [K]$ and we denote c_i the community of node i . Considering the symmetric connectivity matrix $Q \in [0, 1]^{K \times K}$, the adjacency matrix of the graph $X \in \{0, 1\}^{n \times n}$ related to the assignment of the communities $(c_i)_{i \in [n]}$ is defined by

$$X_{i,j} \sim \text{Ber}(Q_{c_i, c_j}),$$

where $\text{Ber}(p)$ indicates a Bernoulli random variable with parameter $p \in [0, 1]$.

For a parameter $\alpha_n \in (0, 1)$ varying with the number of nodes n , we will be focused on connectivity matrix of the form

$$Q := \alpha_n Q_0,$$

where $Q_0 \in [0, 1]^{K \times K}$. As highlighted for example in [3], the rate of α_n as $n \rightarrow \infty$ is a key property to study random graphs sampled from SBMs. Typical regimes are $\alpha_n \sim 1$ (dense regime), $\alpha_n \sim \frac{\log(n)}{n}$ (relatively sparse regime) and $\alpha_n \sim \frac{1}{n}$ (sparse regime).

In the standard Stochastic Block Model, the communities $(c_i)_{i \in [n]}$ are assigned independently to each node according to a probability distribution $\nu \in [0, 1]^K$, $\sum_{k \in [K]} \nu_k = 1$. Stated otherwise, the community c_i of node $i \in [n]$ is randomly sampled from the distribution ν .

Markovian assignment of communities in the SBM We introduce in this paper a new Stochastic Block Model by assigning a community to each node using a Markovian dynamic. We start by ordering the n nodes in V and without loss of generality, we consider the increasing order of the integers $1, 2, \dots, n$. For all $i \in [n]$, we denote $C_i \in [K]$ the random variable representing the community of the node i and we consider that they satisfy the following assumption.

Assumption A1. $(C_i)_{i \in [n]}$ is a positive recurrent Markov chain on the finite space $[K]$ with invariant probability π , with transition matrix $P \in \mathbb{R}^{K \times K}$ and initial distribution π .

We assign communities as follows:

$$\begin{aligned} & C_1 \sim \pi \\ & \text{For } i = 1 \dots (n-1) \text{ Do} \\ & \quad C_{i+1} \sim P_{C_i, :} \\ & \text{EndFor.} \end{aligned}$$

Once the community of each node is assigned, we draw an edge between the nodes i and j with probability Q_{C_i, C_j}

$$X_{i,j} \sim \text{Ber}(Q_{C_i, C_j}) \quad \text{with} \quad Q := \alpha_n Q_0.$$

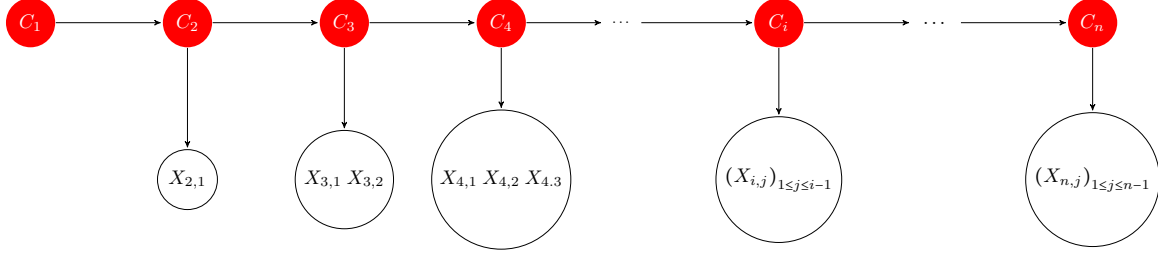


Figure 1: Graphical model presenting the SBM with Markovian assignment of the communities.

Here, $Q_0 \in [0, 1]^{K \times K}$ and $\alpha_n \in (0, 1)$ is varying with n . Figure 1 presents a graphical representation of our model.

We will need the following additional notations

$$L := \|Q_0\|_\infty \quad \pi_m := \min_{c \in [K]} \pi(c) \quad D^2 := \min_{l \neq k} \|(Q_0)_{:,k} - (Q_0)_{:,l}\|^2,$$

and we adopt the following condition which is necessary for the identifiability of our model.

Assumption A2. $D^2 > 0$.

Identifiability If assumptions A1 and A2 hold and if $\alpha_n \log(n) \leq 1/L$, Theorems 2, 3 and 4 prove that we are able to get consistent estimation of the parameters P , π and Q of our model as soon as

$$\alpha_n = \Omega\left(\frac{\log(n)}{n}\right).$$

Stated otherwise, under assumptions A1, A2 and $\alpha_n \log(n) \leq 1/L$, our model is proved to be identifiable as soon as the average degree of the nodes is of order $\log(n)$ or higher.

Error measure Given two partitions $\hat{G} = (\hat{G}_1, \dots, \hat{G}_K)$ and $G = (G_1, \dots, G_K)$ of $[n]$ into K non-void groups, we define the proportion of non-matching points

$$\text{err}(\hat{G}, G) = \min_{\sigma \in \mathcal{S}_K} \frac{1}{2n} \sum_{k=1}^K |\hat{G}_k \Delta G_{\sigma(k)}|,$$

where $A \Delta B$ represents the symmetric difference between the two sets A and B and \mathcal{S}_K represents the set of permutations on $\{1, \dots, K\}$. When \hat{G} is a partition estimating G , we refer to $\text{err}(\hat{G}, G)$ as the misclassification proportion (or error) of the clustering.

The quantity

$$n\alpha_n\pi_mD^2/L$$

will be interpreted as the signal-to-noise ratio to control the misclassification error of our algorithm.

3 Inferring the hidden communities

In this Section, we present how we estimate the partition of the nodes \hat{G} when communities are assigned using a Markovian dynamic. Our main result Theorem 1 shows that we are able to achieve

$$-\log \text{err}(\hat{G}, G) = \Omega(n\alpha_n).$$

Stated otherwise, we get a misclassification error that decays exponentially fast with respect to $n\alpha_n$. We recover the convergence rate recently proved in [11] in the standard SBM² when the size of the smallest cluster scales linearly with n like in our case (see (8) for a proof). To reach this result, we use the SDP algorithm proposed by Giraud and Verzelen in [11]. In the following, we expose how the method works.

²See Theorem B.

Suppose the community of each node in the graph has been assigned. In all this subsection, all the communities are considered fixed. The size of the community $k \in [K]$ will be denoted m_k . The size of the smallest community will be denoted m and X will define the adjacency matrix of the graph.

In [11], the authors are interested in solving optimization problem similar to the following

$$\max_{B \in \mathcal{C}'} \langle X, B \rangle \quad \text{with } \mathcal{C}' := \{B : \text{PSD}, B_{k,l} \geq 0, |B|_1 = \sum_k m_k^2\}, \quad (1)$$

where PSD means that B is positive semidefinite.

We remind that, dealing with two communities, when the values of the probability matrix Q are a constant p on the diagonal and another constant q off the diagonal with $p > q$, we are in the assortative case. In the assortative setting, optimization problems like (1) have been widely used to recover communities, see [6], [12], [23], [13], [9]. Those SDP programs are trying to maximize the probability of connection between nodes belonging to the same community. Therefore, they cannot be used directly to solve community detection outside of the assortative framework.

Peng and Wei in [21] showed that any partition G of $[n]$ can be uniquely represented by a $n \times n$ matrix $B^* \in \mathbb{R}^{n \times n}$ defined by

$$\forall i, j \in [n], \quad B_{i,j}^* = \begin{cases} \frac{1}{m_k} & \text{if } i \text{ and } j \text{ belong to the same community } k \in [K] \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The set of such matrices B^* that can be built from a particular partition of $[n]$ in K groups is defined by

$$\mathcal{S} = \{B \in \mathbb{R}^{n \times n} : \text{symmetric}, B^2 = B, \text{Tr}(B) = K, B\mathbf{1} = \mathbf{1}, B \geq 0\},$$

where $\mathbf{1} \in \mathbb{R}^n$ is the n -dimensional vector with all entries equal to one and where $B \geq 0$ means that all entries of B are nonnegative. Peng and Wei [21] proved that solving the K -means problem

$$\text{Crit}(G) = \sum_{k=1}^K \sum_{i \in G_k} \left\| X_{:,i} - \frac{1}{|G_k|} \sum_{j \in G_k} X_{:,j} \right\|^2,$$

is equivalent to

$$\max_{B \in \mathcal{S}} \langle XX^\top, B \rangle. \quad (3)$$

Writing B^* an optimal solution of (3), an optimal solution for the K -means problem is obtained by gathering indices $i, j \in [n]$ such that $B_{i,j}^* \neq 0$. The set \mathcal{S} is not convex and the authors of [11] propose the following relaxation of problem (3)

$$\hat{B} \in \arg \max_{B \in \mathcal{C}_\beta} \langle XX^\top, B \rangle, \quad (4)$$

where

$$\mathcal{C}_\beta := \{B \in \mathbb{R}^{n \times n} : \text{symmetric}, \text{Tr}(B) = K, B\mathbf{1} = \mathbf{1}, 0 \leq B \leq \beta\} \quad \text{with} \quad K/n \leq \beta \leq 1.$$

The constraints $B \leq \beta$ allows to deal with sparse graphs. Indeed, when $\alpha_n = o(\log(n)/n)$, solving (4) without this constraint will produce unbalanced partition.

At this step, we cannot ensure that \hat{B} belongs to \mathcal{S} and a final refinement is necessary to end up with a clustering of the nodes of the graph. This final rounding step is achieved by running a K -medoid algorithm on the rows of \hat{B} . Given a partition $\{G_1, \dots, G_k\}$ of the n nodes of the graph into K communities, we define the related membership matrix $A \in \mathbb{R}^{n \times K}$ where $A_{i,k} = \mathbb{1}_{i \in G_k}$.

Working on the rows of \hat{B} , a K -medoid algorithm tries to find efficiently a pair (\hat{A}, \hat{M}) with $\hat{A} \in \mathcal{A}_K$, $\hat{M} \in \mathbb{R}^{K \times n}$, $\text{Rows}(\hat{M}) \subset \text{Rows}(\hat{B})$ satisfying for some $\rho > 0$

$$|\hat{A}\hat{M} - \hat{B}|_1 \leq \rho \quad \min_{A \in \mathcal{A}_K, \text{Rows}(M) \subset \text{Rows}(\hat{B})} |AM - \hat{B}|_1, \quad (5)$$

where \mathcal{A}_K is the set of all possible membership matrices and $\text{Rows}(\hat{B})$ the set of all rows of \hat{B} .

The K -medoids algorithm proposed in [5] gives in polynomial time a pair (\hat{A}, \hat{M}) satisfying the inequality (5) with $\rho = 7$. From \hat{A} we are able to define the final partition of the nodes of the graph by setting

$$\forall k \in [K], \quad \hat{G}_k = \{i \in [n] : \hat{A}_{i,k} = 1\}.$$

Remark.

As highlighted in [11], the parameter β can not be computed since L is unknown. Verzelen and Giraud propose to set β to value $\hat{\beta} = \frac{K^3}{n} e^{2nd_X} \wedge 1$, where d_X denotes the density of the graph.

We end up with the Algorithm 1 to estimate the communities in the SBM.

Algorithm 1 Algorithm to estimate the partition of the nodes of the graph.

Data: Adjacency matrix X of a graph $G = (V, E)$, Number of communities K .

- 1: Compute the density of the graph $d_X = \frac{2|E|}{n(n-1)}$ and set $\hat{\beta} = \frac{K^3}{n} e^{2nd_X} \wedge 1$.
 - 2: Find $\hat{B} \in \arg \max_{B \in \mathcal{C}_{\hat{\beta}}} \langle XX^\top, B \rangle$ (using for example the interior-point method).
 - 3: Run the K -medoids algorithm from [5] on the rows of \hat{B} . Note $\hat{A} \in \{0, 1\}^{n \times K}$ the membership matrix obtained.
 - 4: Define $\forall k \in [K]$, $\hat{G}_k = \{i \in [n] : \hat{A}_{i,k} = 1\}$ and $\forall i \in [n]$, $\hat{C}_i = k$ where $k \in [K]$ is such that $\hat{A}_{i,k} = 1$.
-

4 Inferring the parameters of the model

In the following, $(\hat{C}_i)_{1 \leq i \leq n}$ and $(\hat{G}_k)_{k \in [K]}$ denote respectively the estimators of $(C_i)_{1 \leq i \leq n}$ and $(G_k)_{k \in [K]}$ provided by the Algorithm 1.

4.1 The connectivity matrix

Once the partition of the nodes is correctly recovered, a natural estimator for $Q_{k,l}$ (for $k, l \in [K]^2$) consists in computing the ratio between the number of edges between nodes with communities k and l and the maximum number of edges between nodes with communities k and l .

$$\forall k, l \in [K]^2, \hat{Q}_{k,l} := \begin{cases} \frac{1}{|\hat{G}_k| \times |\hat{G}_l|} \sum_{i \in \hat{G}_k} \sum_{j \in \hat{G}_l} X_{i,j} & \text{if } k \neq l \\ \frac{1}{|\hat{G}_k| \times (|\hat{G}_k| - 1)} \sum_{i,j \in \hat{G}_k} X_{i,j} & \text{if } k = l \end{cases}$$

One can remark that each entry of our estimator \hat{Q} is a sum of identically distributed and independent Bernoulli random variables (i.e. it is a Binomial random variable).

4.2 The invariant distribution of the Markov chain

Thanks to the ergodic theorem, we know that the average number of visits of each state of the chain converges toward the invariant probability of the chain at this particular state. Stated otherwise, for all community $k \in [K]$, the average number of nodes with community k in the graph converges toward $\pi(k)$ as n tends to $+\infty$. Therefore we propose to estimate the invariant measure of the chain $(C_i)_{i \geq 1}$ with $\hat{\pi}$ defined by

$$\forall k \in [K], \hat{\pi}_k := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\hat{C}_i=k}.$$

4.3 The transition matrix of the Markov chain

We define $(Y_i)_{i \geq 1}$ a Markov Chain on $[K]^2$ by setting : $Y_i = (C_i, C_{i+1})$. We define naturally the sequence $(\hat{Y}_i)_{i \geq 1}$ by $\hat{Y}_i = (\hat{C}_i, \hat{C}_{i+1})$. The transition kernel of the Markov Chain $(Y_i)_{i \geq 1}$ is $\mathcal{P}_{(k,l),(k',l')} = \mathbb{1}_{l=k'} P_{k,l}$ and its invariant measure is given by μ such that $\forall k, l, \mu(k, l) = \pi(k) P_{k,l}$. We propose to estimate each entry of the transition matrix P of the Markov chain $(C_i)_{i \geq 1}$ with

$$\forall k, l \in [K]^2, \hat{P}_{k,l} := \frac{n}{n-1} \frac{\sum_{i=1}^{n-1} \mathbb{1}_{\hat{Y}_i=(k,l)}}{\sum_{i=1}^n \mathbb{1}_{\hat{C}_i=k}}.$$

5 Results

We provide a partial recovery bound in the Stochastic Block Model when the communities are assigned through a Markovian dynamic. We define the signal-to-noise ratio as

$$S^2 := \frac{n\alpha_n\pi_m D^2}{L},$$

reminding that $\pi_m = \min_{c \in [K]} \pi(c)$, $\|Q_0\|_\infty \leq L$ and $D^2 = \min_{l \neq k} \|(Q_0)_{:,k} - (Q_0)_{:,l}\|^2$. We shed light on the fact that this quantity matches asymptotically the SNR from Theorem B when π is the uniform distribution over $[K]$ and when the communities are assigned independently to each node according to the probability distribution π . Similarly to Theorem B, we prove with Theorem 1 that the misclassification error decays exponentially fast with respect to the SNR S^2 .

In Theorems 1, 2, 3 and 4, the constants a and b' only depend on the parameters π , P and Q_0 while the constant b also depends on the number of communities K . In Section 7, Lemmas 1.1, 2, 3 and 4 provide respectively a more complete version of Theorems 1, 2, 3 and 4 by giving explicitly these constants.

Theorem 1. *Assume that $\alpha_n \log(n) \leq 1/L$. Then there exist three constants $a, b, c > 0$ such that for any n satisfying*

$$n\alpha_n > a,$$

it holds with probability at least $1 - b/n^2$

$$\text{err}(\hat{G}, G) \leq e^{-cS^2}.$$

In particular, it holds with probability at least $1 - b/n^2$

$$-\log(\text{err}(\hat{G}, G)) = \Omega(n\alpha_n).$$

Remark. Theorem 1 states that in the relatively sparse setting (i.e. when $\alpha_n \sim \log(n)/n$), we achieve a polynomial decay of the misclassification error with order $\pi_m D^2/L$. The greater the quantity $\pi_m D^2/L$ is, the faster the misclassification error decays. The condition on the sparsity parameter α_n indicates that we are able to deal with the sparse regime (i.e. when $\alpha_n \sim 1/n$) for n large enough when $\lim_{n \rightarrow \infty} n\alpha_n > a$.³

In Theorems 2, 3 and 4, the condition on the sparsity parameter α_n indicates that we get consistent estimation respectively of the transition matrix, the invariant measure and the connectivity matrix in the relatively sparse regime (i.e. when $\alpha_n \sim \log(n)/n$) for n large enough when $\lim_{n \rightarrow \infty} n\alpha_n/\log(n) > a$.

Theorem 2. *Let us consider $\gamma > \frac{5K}{2\pi_m^2}$.*

Assume that $\alpha_n \log(n) \leq 1/L$. Then there exist three constants $a, b, b' > 0$ such that for any n satisfying

$$\frac{n\alpha_n}{\log(n)} \geq a \quad \text{and} \quad n\alpha_n \geq \frac{a}{\gamma^2},$$

it holds with probability at least $1 - b \left[1/n^2 \vee \exp\left(-b'(\gamma - \frac{5K}{2\pi_m^2})^2\right) \right]$,

$$\|\hat{P} - P\|_\infty \leq \frac{\gamma}{\sqrt{n}}.$$

Remark.

To prove this theorem, we consider the Markov chain $(Y_i)_{i \geq 1}$ built considering two consecutive states of the Markov chain $(C_i)_{i \geq 1}$. Stated otherwise, the state number i of the Markov chain used is formed by the couple of the communities of the nodes number i and number $i + 1$.

Theorem 3. *Let us consider $\gamma > 0$.*

Assume that $\alpha_n \log(n) \leq 1/L$. Then there exist three constants $a, b, b' > 0$ such that for any n satisfying

$$\frac{n\alpha_n}{\log(n)} \geq a,$$

it holds with probability at least $1 - b(1/n^2 \vee \exp(-b'\gamma^2))$

$$\|\hat{\pi} - \pi\|_\infty \leq \frac{\gamma}{\sqrt{n}}.$$

³Note that the constants a in Theorems 2, 3 and 4 are different. See Lemmas 2, 3 and 4 to get explicitly these constants.

Theorem 4. *Let us consider $\gamma > 0$.*

Assume that $\alpha_n \log(n) \leq 1/L$. Then there exist three constants $a, b, b' > 0$ such that for any n satisfying

$$\frac{n\alpha_n}{\log(n)} \geq a \quad \text{and} \quad n > \left(\frac{\gamma + 1}{\pi_m} \right)^2,$$

it holds with probability at least $1 - b(1/n^2 \vee \exp(-b'\gamma^2))$

$$\|\hat{Q} - Q\|_\infty \leq \frac{\gamma}{\sqrt{n}}.$$

6 Simulations

The code to reproduce the following results is available [here](#)⁴.

6.1 Experiments with 2 communities

We test our algorithm on a toy example with $K = 2$ communities, $\alpha_n = 1$ and with the following matrices:

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix} \text{ and } Q_0 = \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.3 \end{bmatrix}. \quad (6)$$

The Figure 2 shows the evolution of the infinity norm of the difference between the true transition matrix P and our estimate \hat{P} when the size of the graph is increasing. Those numerical results are consistent with Theorem 2 : we recover the parametric convergence rate with our estimator of the transition matrix.

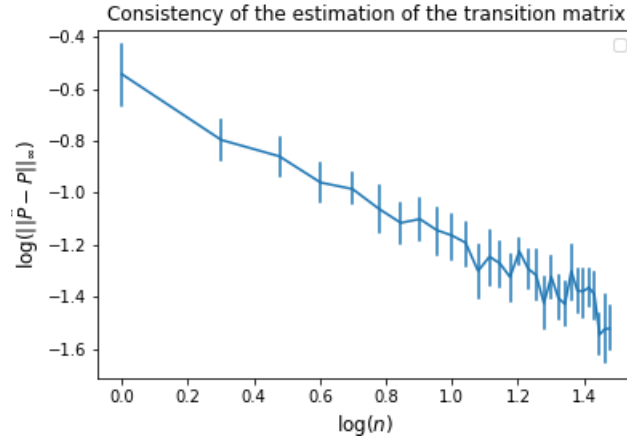


Figure 2: We plot the log of the infinity norm of the difference between the true transition matrix P and our estimate \hat{P} according to the log of the number of nodes in the graph. For each point, the bar represents the standard deviation of the infinity norm error computed over thirty randomly generated graphs with the same number of nodes and using the matrices P and Q defined by (6).

With Figure 3, we shed light on the influence of the average degree of the nodes on the performance of our algorithm. We propose to compute the precision and the recall of the binary classification problem that we study when $K = 2$ defining

$$Q = \alpha \times Q_0,$$

where Q_0 is defined in (6) and α varies between 0.1 and 1 on a log scale. We remind that in a binary classification problem, the precision is the ratio between the number of examples labeled 1 that belong to class 1 and the number of examples labeled 1. The recall is the ratio between the number of examples

⁴<https://github.com/quentin-duchemin/inference-markovian-SBM>

labeled 1 that belong to class 1 and the number of examples that belong to class 1. In our context, those definitions read as

$$\text{precision} = \frac{\sum_{i=1}^n \mathbb{1}\{\hat{C}_i = 1, C_i = 1\}}{\sum_{i=1}^n \mathbb{1}\{\hat{C}_i = 1\}} \quad \text{and} \quad \text{recall} = \frac{\sum_{i=1}^n \mathbb{1}\{\hat{C}_i = 1, C_i = 1\}}{\sum_{i=1}^n \mathbb{1}\{C_i = 1\}}.$$

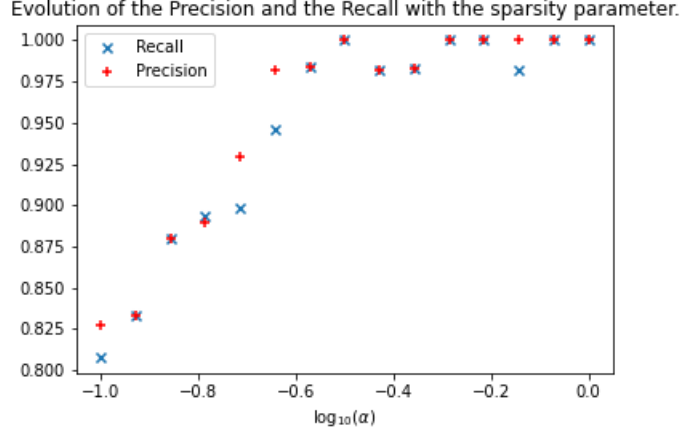


Figure 3: We plot the recall and the precision of the output of our algorithm with a graph sampled from SBM with a Markovian assignment of the communities using $n = 100$ nodes, a transition matrix P defined in (6) and a connectivity matrix $Q = \alpha Q_0$ where Q_0 is defined in (6) and α varies on a log scale between 0.1 and 1. We show the recall and the precision with respect to the \log_{10} of the sparsity parameter α .

6.2 Experiments with 5 communities

We test our algorithm on a toy example with $K = 5$ communities, with the transition matrix P and the connectivity matrix Q defined by

$$P = \begin{bmatrix} 0.1 & 0.3 & 0.5 & 0.01 & 0.09 \\ 0.55 & 0.15 & 0.1 & 0.05 & 0.15 \\ 0.15 & 0.3 & 0.1 & 0.2 & 0.25 \\ 0.15 & 0.05 & 0.1 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.1 & 0.05 & 0.35 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} 0.6 & 0.1 & 0.15 & 0.1 & 0.2 \\ 0.2 & 0.5 & 0.35 & 0.1 & 0.4 \\ 0.4 & 0.15 & 0.6 & 0.25 & 0.05 \\ 0.4 & 0.1 & 0.1 & 0.2 & 0.55 \\ 0.3 & 0.35 & 0.2 & 0.1 & 0.7 \end{bmatrix}. \quad (7)$$

Sampling random graphs from SBM with Markovian assignment of the communities using the matrices (7), we see with Figure 4 that communities 3 and 4 have small sizes compared to the other clusters. For a graph sampled with a size equal to 40, Figure 4.a shows us that the SDP algorithm defined in 1 is able to capture relevant information about the clustering of the nodes in communities 1, 2 and 5. However, we see that using a number of nodes equal to 40 is not enough to distinguish nodes belonging to community 3 or 4. Figure 4.b proves that increasing the size of the graph (with $n = 160$) allows to solve this issue. One can easily guess that running a K -medoid algorithm on the rows of the matrix \hat{B} plotted in Figure 4.b will lead to an accurate clustering of the nodes of the graph. Figure 5 shows that the log of the misclassification error decreases linearly with the size of the graph.

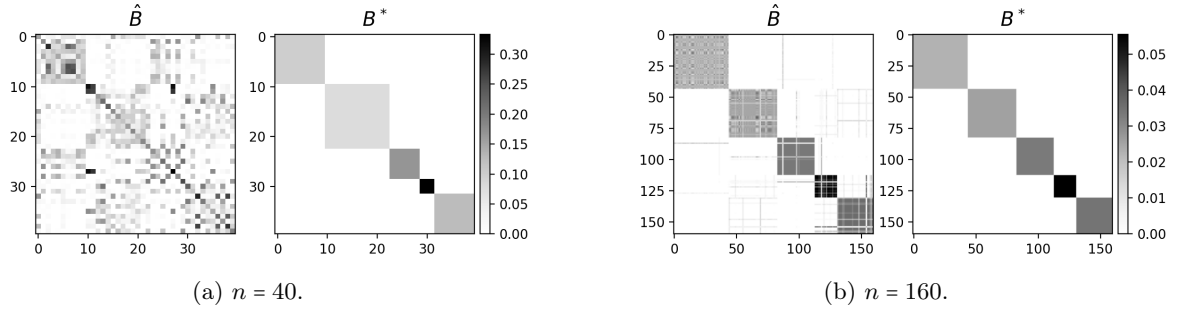


Figure 4: We consider $K = 5$ communities and we order the nodes of the graph such that the true partition of the nodes is given by $G_1 = \{1, \dots, m_1\}$, $G_2 = \{m_1 + 1, \dots, m_1 + m_2\}$, \dots , $G_5 = \{\sum_{j=1}^4 m_j + 1, \dots, n\}$. We generate random graphs from SBM with Markovian assignment of the communities using the transition matrix P and the connectivity matrix Q defined by 7. We plot the matrix B^* solution of (3) and its approximation \hat{B} obtained by solving the SDP (4). Thanks to the node ordering, the matrix B^* (defined in (2)) has a block diagonal structure where each entry of one block is equal to the inverse of the size of the associated cluster. Figure (a) allows us to compare the matrices B^* and \hat{B} when the number of nodes in the graph is equal to 40 while Figure (b) deals with a graph of size 160.

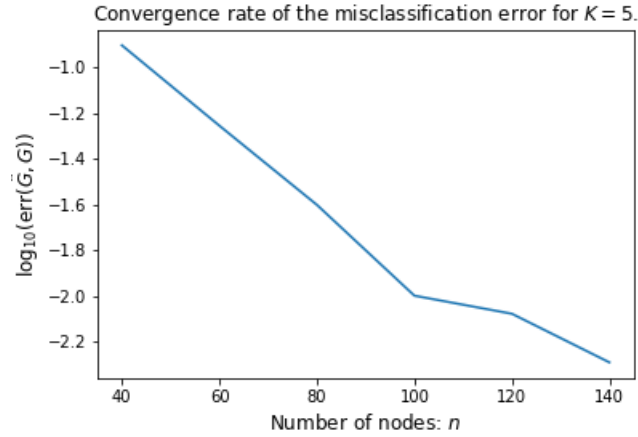


Figure 5: We consider $K = 5$ communities and we sample random graph from SBM with Markovian assignment of the communities using matrices defined in (7). We estimate the partition of the nodes of the graph using Algorithm 1. We plot the log of the misclassification error as a function of the size of the graphs sampled.

7 Proofs

7.1 Proof of Theorem 1

Lemma 1.1 provides a more complete version of Theorem 1 by giving explicitly the constants.

Lemma 1.1. Let us consider the three positive constants c , c' and c'' involved in Theorem B.

Assume that $\alpha_n \log(n) \leq 1/L$ and that $n\alpha_n > \max\left(\frac{4Lc''}{\pi_m^2 D^2}, \frac{2}{L\pi_m}\right)$. Then it holds

$$\mathbb{P}\left(\text{err}(\hat{G}, G) > \exp\left(-\frac{c'S^2}{2}\right)\right) \leq \frac{c}{n^2} + 2K \exp\left(-\frac{n\pi_m^2}{2A_1 + 4A_2\pi_m}\right),$$

where $S^2 = \frac{n\alpha_n\pi_mD^2}{L}$ and where A_1 and A_2 are constants that only depend on the Markov chain $(C_i)_{i \geq 1}$ with $A_1 := \frac{1 + (\lambda_+ \vee 0)}{1 - (\lambda_+ \vee 0)}$ and $A_2 := \frac{1}{3}\mathbb{1}_{\lambda_+ \leq 0} + \frac{5}{1 - \lambda_+}\mathbb{1}_{\lambda_+ > 0}$. Here $1 - \lambda_+$ is the right \mathcal{L}_2 spectral gap of the Markov chain $(C_i)_{i \geq 1}$ (see Definition 1 in Appendix A).

Remarks.

- The fact that $\pi_m > 0$ is a direct consequence of the positive recurrent property of the Markov chain.
- The second term in the right hand side of the previous inequality comes from the concentration of the average number of visits of the Markov chain towards the invariant distribution of the chain. The first term in this inequality corresponds to the bound from Theorem B when communities have been assigned. Recalling that $\|Q\|_\infty$ is upper bounded by $\alpha_n L$, the condition $\alpha_n \log(n) \leq 1/L$ enforces the signal to noise ratio defined by Giraud and Verzelen $s^2 := \Delta^2/(\alpha_n L)$ (see Theorem B) to be larger than $\Delta^2 \times \log(n)$. Another way to interpret this condition is to say that it enforces the expected degree of all nodes of the graph to be smaller than $n/\log(n)$.

- In order to get some intuition on the conditions on n in the previous theorem, keep in mind that asymptotically, the size of the smallest community in the graph will be $n \times \pi_m$.

- The condition $n > \frac{4Lc''}{\alpha_n \pi_m^2 D^2}$ can be read as $(n \times \pi_m) \alpha_n D^2 / L > \frac{4c''}{\pi_m} = 4c'' \times \frac{n}{n\pi_m}$. Asymptotically, $(n \times \pi_m) \alpha_n D^2 / L$ provides a lower bound on the signal-to-noise ratio defined in Theorem B. This shows that the condition $n > \frac{4Lc''}{\alpha_n \pi_m^2 D^2}$ is related to the constraint $s^2 \gtrsim n/m$ of Theorem B.
- The condition $n > \frac{2}{\alpha_n L \pi_m}$ can be read as $\frac{1}{n \times \pi_m} < \alpha_n L / 2$. This shows that the condition $n > \frac{2}{\alpha_n L \pi_m}$ is related to the constraint $1/m < \alpha_n L$ from Theorem B.

The proof of Lemma 1.1 is based on the following Lemma which is proved [below](#).

Lemma 1.2. We consider c, c' and c'' the three numerical constants involved in Theorem B.

Let us consider $0 < t < \pi_m$. Assume that $\alpha_n L \leq 1/\log(n)$. Then for any $\epsilon > 0$ and n large enough such that:

$$n \times (\pi_m - t) \geq \begin{cases} \frac{L \log(1/\epsilon)}{c' \alpha_n D^2} & (i) \\ \left(\frac{c'' n L}{\alpha_n D^2} \right)^{1/2} & (ii) \\ 1/(\alpha_n L) & (iii) \end{cases}$$

it holds

$$\mathbb{P}(\text{err}(\hat{G}, G) > \epsilon) \leq \frac{c}{n^2} + 2K \exp\left(-\frac{nt^2}{2(A_1/4 + A_2 t)}\right),$$

where A_1 and A_2 are constants defined in Theorem 1.

Note that the only constraint on ϵ is given by the condition (i) which is equivalent to

$$\epsilon \geq \exp\left(-\frac{c' D^2 n \alpha_n (\pi_m - t)}{L}\right).$$

In order to get the tighter result possible, we want to choose $\epsilon = \exp\left(-\frac{c' D^2 n \alpha_n (\pi_m - t)}{L}\right)$ which leads to

$$t = \pi_m - \frac{L \log(1/\epsilon)}{c' D^2 n \alpha_n}.$$

The condition $t > 0$ is then equivalent to

$$\pi_m > \frac{L \log(1/\epsilon)}{c' D^2 n \alpha_n} \Leftrightarrow \exp(-\pi_m n \alpha_n c' D^2 / L) < \epsilon.$$

The condition (ii) is equivalent to

$$n(\pi_m - t) = \frac{L \log(1/\epsilon)}{c' \alpha_n D^2} \geq \left(\frac{c'' n L}{\alpha_n D^2} \right)^{1/2} \Leftrightarrow \exp \left(-c' \sqrt{\frac{D^2 c'' n \alpha_n}{L}} \right) \geq \epsilon.$$

The condition (iii) is equivalent to

$$n(\pi_m - t) = \frac{L \log(1/\epsilon)}{c' \alpha_n D^2} \geq (1/\alpha_n L) \Leftrightarrow \exp \left(-\frac{c' D^2}{L^2} \right) \geq \epsilon.$$

One can easily prove that for $n\alpha_n > \max \left(\frac{4Lc''}{\pi_m^2 D^2}, \frac{2}{L\pi_m} \right)$, $\epsilon := \exp \left(-\frac{\pi_m n \alpha_n c' D^2}{2L} \right)$ satisfies the three conditions above. We proved Lemma 1.1. \blacksquare

Proof of Lemma 1.2. Using Theorem 1.2 from [16], we get that

$$\forall c \in [K], \forall t > 0, \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=c} - \pi(c) \right| \geq t \right) \leq 2 \exp \left(-\frac{nt^2}{2(A_1 \sigma_c^2 + A_2 t)} \right) \quad (8)$$

where $A_1 = \frac{1 + (\lambda_+ \vee 0)}{1 - (\lambda_+ \vee 0)}$, $A_2 = \frac{1}{3} \mathbb{1}_{\lambda_+ \leq 0} + \frac{5}{1 - \lambda_+} \mathbb{1}_{\lambda_+ > 0}$ and $\sigma_c^2 = \pi(c)(1 - \pi(c))$.

We deduce that for all $t > 0$,

$$\mathbb{P} \left(\bigcup_c \left\{ \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=c} - \pi(c) \right| \geq t \right\} \right) \leq 2K \exp \left(-\frac{nt^2}{2(A_1 \sigma^2 + A_2 t)} \right),$$

where $\sigma^2 := \max_c \sigma_c^2$ ($\leq 1/4$).

We define $\Omega^c := \bigcup_c \left\{ \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=c} - \pi(c) \right| \geq t \right\}$ and we recall $\pi_m = \min_c \pi(c)$ and $D^2 = \min_{j \neq k} \sum_l ((Q_0)_{k,l} - (Q_0)_{j,l})^2$.

Suppose that $0 < t < \pi_m$ and that n is large enough to satisfy :

$$n \times (\pi_m - t) \geq \begin{cases} \frac{L \log(1/\epsilon)}{c' \alpha_n D^2} & (i) \\ \left(\frac{c'' n L}{\alpha_n D^2} \right)^{1/2} & (ii) \\ 1/(\alpha_n L) & (iii) \end{cases}$$

$$\begin{aligned} \mathbb{P}(\text{err}(\hat{G}, G) > \epsilon) &= \mathbb{P}(\{\text{err}(\hat{G}, G) > \epsilon\} \cap \Omega) + \mathbb{P}(\{\text{err}(\hat{G}, G) > \epsilon\} \cap \Omega^c) \\ &\leq \mathbb{P}(\{\text{err}(\hat{G}, G) > \epsilon\} \cap \Omega) + 2K \exp \left(-\frac{nt^2}{2(A_1 \sigma^2 + A_2 t)} \right) \\ &= \mathbb{P}(\text{err}(\hat{G}, G) > \epsilon \mid \Omega) \times \mathbb{P}(\Omega) + 2K \exp \left(-\frac{nt^2}{2(A_1 \sigma^2 + A_2 t)} \right). \quad (*) \end{aligned}$$

We denote by M the random variable that gives the size of the smallest cluster: $M := \min_{k \in [K]} m_k$. Condition (i) is equivalent to

$$\epsilon \geq \exp \left(-c' \frac{n \alpha_n (\pi_m - t) D^2}{L} \right).$$

Since on the event Ω , we have $n(\pi_m - t) \leq M$ almost surely, we get that on Ω ,

$$\epsilon \geq \exp \left(-c' \frac{M \alpha_n D^2}{L} \right) \geq \exp(-c' s^2),$$

where $s^2 = \Delta^2/(\alpha_n L)$ with $\Delta^2 = \min_{k \neq j} \Delta_{k,j}^2$ and $\Delta_{k,j}^2 = \sum_l m_l (Q_{k,l} - Q_{j,l})^2$. The last equality comes from that $\Delta^2 \geq M \alpha_n^2 D^2$. Using (*) we get that

$$\begin{aligned}\mathbb{P}(\text{err}(\hat{G}, G) > \epsilon) &\leq \mathbb{P}(\text{err}(\hat{G}, G) > \epsilon \mid \Omega) + 2K \exp\left(-\frac{nt^2}{2(A_1\sigma^2 + A_2t)}\right) \\ &\leq \mathbb{P}(\text{err}(\hat{G}, G) > e^{-c's^2} \mid \Omega) + 2K \exp\left(-\frac{nt^2}{2(A_1\sigma^2 + A_2t)}\right).\end{aligned}$$

We note that on Ω :

- Condition (ii) gives

$$M^2 \geq \frac{c''nL}{\alpha_n D^2} \Leftrightarrow \frac{M\alpha_n D^2}{L} \geq c''n/M,$$

which implies that $s^2 = \frac{\Delta^2}{\alpha_n L} \geq c''n/M$ since $\Delta^2 \geq M\alpha_n^2 D^2$.

- Condition (iii) gives

$$\frac{1}{M} \leq \alpha_n L.$$

Applying the result of Verzelen and Giraud from [11], we get that

$$\mathbb{P}(\text{err}(\hat{G}, G) > e^{-c's^2} \mid \Omega) \leq \frac{c}{n^2}.$$

Finally,

$$\mathbb{P}(\text{err}(\hat{G}, G) > \epsilon) \leq \frac{c}{n^2} + 2K \exp\left(-\frac{nt^2}{2(A_1\sigma^2 + A_2t)}\right).$$

■

7.2 Proof of Theorem 2

We will prove a more accurate result with Lemma 2.

Lemma 2. Let us consider $\gamma > \frac{5K}{2\pi_m^2}$.

Assume that $\alpha_n \log(n) \leq 1/L$, that $n\alpha_n > \max\left(\frac{4Lc''}{\pi_m^2 D^2}, \frac{4}{L\pi_m}, \frac{2L \log(n)}{\pi_m c' D^2}\right)$ and that $\sqrt{n} > \frac{2}{\pi_m}(1 + \pi_m^2 \gamma/5)$. Then it holds

$$\mathbb{P}\left(\|\hat{P} - P\|_\infty \geq \frac{\gamma}{\sqrt{n}}\right) \leq 2K^2 \exp\left(-\frac{\left(\frac{\pi_m^2 \gamma}{5K} - \frac{1}{2}\right)^2}{2(B_1/4 + B_2 \frac{\pi_m^2 \gamma}{5K\sqrt{n}})}\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{n\pi_m^2}{8A_1\sigma^2 + 4A_2\pi_m}\right),$$

where B_1 and B_2 depend only on the Markov chain and are defined by $B_1 := \frac{1 + (\xi_+ \vee 0)}{1 - (\xi_+ \vee 0)}$ and $B_2 := \frac{1}{3}\mathbb{1}_{\xi_+ \leq 0} + \frac{5}{1 - \xi_+}\mathbb{1}_{\xi_+ > 0}$. Here $1 - \xi_+$ is the right \mathcal{L}_2 spectral gap of the Markov chain $(Y_i)_{i \geq 1}$ (see Definition 1 in Appendix A).

Remarks.

- The first term in the right hand side of the previous inequality is due to the concentration of the average number of visits of the chain $(Y_i)_{i \geq 1}$ (defined in subsection 4.3) towards its invariant distribution. The two last terms of the inequality correspond to the bound guaranteeing the recovery of the true partitions with a direct application of Theorem 1.
- The condition $n\alpha_n > \frac{2L \log(n)}{\pi_m c' D^2}$ ensures that $\exp\left(-\frac{\pi_m n \alpha_n c' D^2}{2L}\right) < \frac{1}{n}$. Theorem 1 will then guarantee that we recover perfectly the partition of the communities.

- Expecting the accuracy γ/\sqrt{n} , the condition $\sqrt{n} > \frac{2}{\pi_m}(1 + \pi_m^2 \gamma/5)$ ensures that the Markov chain $(C_i)_{i \geq 1}$ has visited enough each state $k \in [K]$ to guarantee the convergence of the average number of visits toward the invariant distribution.

Proof of Lemma 2.

I. Concentration of the average number of visits for $(Y_i)_{i \geq 1}$.

We recall that $(Y_i)_{i \geq 1}$ is a Markov Chain on $[K]^2$ defined by : $Y_i = (C_i, C_{i+1})$. Then using again Theorem 1.2 from [16], we get that

$$\forall t > 0, \forall k, l \in [K]^2, \mathbb{P}\left(\left|\frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)} - \pi(k)P_{k,l}\right| \geq t\right) \leq 2 \exp\left(-\frac{nt^2}{2(B_1/4 + B_2t)}\right),$$

II. First step toward the theorem.

We define the event $N := \left\{\text{err}(\hat{G}, G) < \exp\left(-\frac{\pi_m n \alpha_n c' D^2}{2L}\right)\right\}$. Note that on N , the partition of the clusters is correctly recovered thanks to the condition $n \alpha_n > \frac{2L \log(n)}{\pi_m c' D^2}$. Let $\gamma > \frac{5K}{2\pi_m^2}$ and let us define

$$r = \frac{\zeta}{\sqrt{n}} \quad \text{with } \zeta = \frac{\pi_m^2 \gamma}{5K} - \frac{1}{2} > 0 \quad \text{and} \quad \Gamma = \bigcap_{k,l} \left\{\left|\frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)} - \pi(k)P_{k,l}\right| < r\right\}.$$

Then,

$$\mathbb{P}\left(\bigcup_{k,l} \left\{|\hat{P}_{k,l} - P_{k,l}| \geq \frac{\gamma}{\sqrt{n}}\right\}\right) \leq \underbrace{\mathbb{P}\left(\bigcup_{k,l} \left\{|\hat{P}_{k,l} - P_{k,l}| \geq \frac{\gamma}{\sqrt{n}}\right\} \mid N, \Gamma\right)}_{(*)} \mathbb{P}(N) \mathbb{P}(\Gamma \mid N) + \mathbb{P}(\Gamma^c) + \mathbb{P}(N^c).$$

Note that the condition $\sqrt{n} > \frac{2}{\pi_m}(1 + \pi_m^2 \gamma/5)$ of Lemma 2 implies

$$\sqrt{n} > \frac{2}{\pi_m}(1 + K\zeta). \tag{9}$$

III. We prove that $(*)$ is zero.

In this third step of the proof, we are going to show that conditionally on the event $N \cap \Gamma$, the infinite norm between our estimate of the transition matrix \hat{P} and P is smaller than γ/\sqrt{n} .

① We split $(*)$ in two terms.

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k,l} \left\{|\hat{P}_{k,l} - P_{k,l}| \geq \frac{\gamma}{\sqrt{n}}\right\} \mid N, \Gamma\right) &= \mathbb{P}\left(\bigcup_{k,l} \left\{\left|\hat{P}_{k,l} - \frac{1}{n-1} \frac{\sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)}}{\pi(k)} + \frac{1}{n-1} \frac{\sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)}}{\pi(k)} - P_{k,l}\right| \geq \frac{\gamma}{\sqrt{n}}\right\} \mid N, \Gamma\right) \\ &\leq \sum_{k,l} \mathbb{P}\left(\left|\hat{P}_{k,l} - \frac{1}{n-1} \frac{\sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)}}{\pi(k)}\right| \geq \frac{\gamma}{2\sqrt{n}} \mid N, \Gamma\right) \\ &\quad + \sum_{k,l} \mathbb{P}\left(\left|\frac{1}{n-1} \frac{\sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)}}{\pi(k)} - P_{k,l}\right| \geq \frac{\gamma}{2\sqrt{n}} \mid N, \Gamma\right) \end{aligned}$$

② We show that on Γ : $\left|\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=k} - \pi(k)\right| \leq \frac{1}{n} + Kr$.

Here we show that a concentration of the average number of visits for $(Y_i)_{i \geq 1}$ gives for free a concentration result of the average number of visits for $(C_i)_{i \geq 1}$.

Note that on the event Γ :

$$\begin{aligned}
& \bullet \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=k} \geq \frac{1}{n} \sum_{l=1}^K \sum_{i=1}^{n-1} \mathbb{1}_{C_i=k, C_{i+1}=l} \\
& = \frac{n-1}{n} \sum_{l=1}^K \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{1}_{C_i=k, C_{i+1}=l} \geq \frac{n-1}{n} \sum_{l=1}^K (\pi(k)P_{k,l} - r) = \frac{n-1}{n} (\pi(k) - Kr).
\end{aligned}$$

$$\text{Hence } \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=k} - \pi(k) \geq -\frac{\pi(k)}{n} - \frac{n-1}{n} Kr \geq -\left(\frac{1}{n} + Kr\right).$$

$$\begin{aligned}
& \bullet \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=k} \leq \frac{1}{n} \sum_{l=1}^K \sum_{i=1}^{n-1} \mathbb{1}_{C_i=k, C_{i+1}=l} + \frac{1}{n} \\
& = \frac{n-1}{n} \sum_{l=1}^K \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{1}_{C_i=k, C_{i+1}=l} + \frac{1}{n} \leq \frac{n-1}{n} \sum_{l=1}^K (\pi(k)P_{k,l} + r) + \frac{1}{n} \leq \pi(k) + Kr + \frac{1}{n}
\end{aligned}$$

$$\text{Hence } \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=k} - \pi(k) \leq \frac{1}{n} + Kr.$$

We deduce then that on Γ , $\left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=k} - \pi(k) \right| \leq \frac{1}{n} + Kr$.

③ We show that the first term from ① is zero.

$$\begin{aligned}
\mathbb{P} \left(\left| \hat{P}_{k,l} - \frac{1}{n-1} \frac{\sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)}}{\pi(k)} \right| \geq \frac{\gamma}{2\sqrt{n}} \mid N, \Gamma \right) &= \mathbb{P} \left(\frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)} \left| \frac{n}{\sum_{i=1}^n \mathbb{1}_{C_i=k}} - \frac{1}{\pi(k)} \right| \geq \frac{\gamma}{2\sqrt{n}} \mid N, \Gamma \right) \\
&\leq \mathbb{P} \left((r + \pi(k)P_{k,l}) \left| \frac{n}{\sum_{i=1}^n \mathbb{1}_{C_i=k}} - \frac{1}{\pi(k)} \right| \geq \frac{\gamma}{2\sqrt{n}} \mid N, \Gamma \right) \\
&= \mathbb{P} \left(\left| \frac{n\pi(k) - \sum_{i=1}^n \mathbb{1}_{C_i=k}}{\pi(k) \sum_{i=1}^n \mathbb{1}_{C_i=k}} \right| \geq \frac{1}{2r + 2\pi(k)P_{k,l}} \cdot \frac{\gamma}{\sqrt{n}} \mid N, \Gamma \right) \text{ and using ②,} \\
&\leq \mathbb{P} \left(\frac{\frac{1}{n} + Kr}{\pi(k)(\pi(k) - \frac{1}{n} - Kr)} \geq \frac{1}{2r + 2\pi(k)P_{k,l}} \cdot \frac{\gamma}{\sqrt{n}} \mid N, \Gamma \right) \\
&\leq \mathbb{P} \left(\frac{\frac{1}{n} + Kr}{\pi_m(\pi_m - \frac{1}{n} - Kr)} \geq \frac{1}{2r + 2} \cdot \frac{\gamma}{\sqrt{n}} \mid N, \Gamma \right) \text{ and since } r = \frac{\zeta}{\sqrt{n}}, \\
&= \mathbb{P} \left(\frac{\frac{1}{n} + K \frac{\zeta}{\sqrt{n}}}{\pi_m(\pi_m - \frac{1}{n} - K \frac{\zeta}{\sqrt{n}})} \geq \frac{1}{2 \frac{\zeta}{\sqrt{n}} + 2} \cdot \frac{\gamma}{\sqrt{n}} \mid N, \Gamma \right) \\
&= \mathbb{P} \left(\frac{(\frac{1}{n} + K \frac{\zeta}{\sqrt{n}})(2\zeta + 2\sqrt{n})}{\pi_m(\pi_m - \frac{1}{n} - K \frac{\zeta}{\sqrt{n}})} \geq \gamma \mid N, \Gamma \right) \\
&\leq \mathbb{P} \left(\frac{(\frac{1}{n} + K \frac{\zeta}{\sqrt{n}})(2\zeta + 2\sqrt{n})}{\pi_m(\pi_m - \frac{1}{\sqrt{n}}(1 + K\zeta))} \geq \gamma \mid N, \Gamma \right). \tag{10}
\end{aligned}$$

Since from (9), $\sqrt{n} \geq \frac{2}{\pi_m}(1 + K\zeta)$, we have

$$\frac{\pi_m}{2} \leq \pi_m - \frac{1}{\sqrt{n}}(1 + K\zeta),$$

which leads to

$$\mathbb{P} \left(\frac{(\frac{1}{n} + K \frac{\zeta}{\sqrt{n}})(2\zeta + 2\sqrt{n})}{\pi_m(\pi_m - \frac{1}{\sqrt{n}}(1 + K\zeta))} \geq \gamma \mid N, \Gamma \right) \leq \mathbb{P} \left(\frac{2(\frac{1}{\sqrt{n}} + K\zeta)(\frac{\zeta}{\sqrt{n}} + 1)}{\pi_m^2/2} \geq \gamma \mid N, \Gamma \right).$$

Moreover, since from (9) and the fact that $\pi_m \in (0, 1)$, $\sqrt{n} \geq \frac{2}{\pi_m}(1 + K\zeta) > 2K\zeta$, it holds

$$\frac{\zeta}{\sqrt{n}} < \frac{1}{2K} < \frac{1}{4}.$$

Coming back to (10), we finally get

$$\begin{aligned} \mathbb{P}\left(\left|\hat{P}_{k,l} - \frac{1}{n-1} \frac{\sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)}}{\pi(k)}\right| \geq \frac{\gamma}{2\sqrt{n}} \mid N, \Gamma\right) &\leq \mathbb{P}\left(\frac{2(\frac{1}{\sqrt{n}} + K\zeta)(\frac{\zeta}{\sqrt{n}} + 1)}{\pi_m^2/2} \geq \gamma \mid N, \Gamma\right) \\ &\leq \mathbb{P}\left(\frac{5(\frac{1}{\sqrt{n}} + K\zeta)}{\pi_m^2} \geq \gamma \mid N, \Gamma\right) \\ &= 0. \end{aligned}$$

The last equality is due to the definition of ζ . Indeed,

$$\zeta = \frac{\gamma\pi_m^2}{5K} - \frac{1}{2} < \frac{\gamma\pi_m^2}{5K} - \frac{1}{K\sqrt{n}} \quad \text{leading to} \quad \frac{5(\frac{1}{\sqrt{n}} + K\zeta)}{\pi_m^2} < \gamma.$$

④ We show that the second term from ① is zero.

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n-1} \frac{\sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)}}{\pi(k)} - P_{k,l}\right| \geq \frac{\gamma}{2\sqrt{n}} \mid N, \Gamma\right) &= \mathbb{P}\left(\left|\frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)} - \pi(k)P_{k,l}\right| \geq \pi(k)\frac{\gamma}{2\sqrt{n}} \mid N, \Gamma\right) \\ &\leq \mathbb{P}\left(\left|\frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{1}_{Y_i=(k,l)} - \pi(k)P_{k,l}\right| \geq \pi_m\frac{\gamma}{2\sqrt{n}} \mid N, \Gamma\right) \\ &= 0 \end{aligned}$$

where the last equality comes from the definition of $r = \zeta/\sqrt{n}$ and the definition of Γ because

$$\zeta = \frac{\pi_m^2\gamma}{5K} - \frac{1}{2} < \frac{\pi_m^2\gamma}{5K} \leq \frac{\pi_m\gamma}{2}.$$

IV. Conclusion.

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k,l} \left\{|\hat{P}_{k,l} - P_{k,l}| \geq \frac{\gamma}{\sqrt{n}}\right\}\right) &\leq \mathbb{P}\left(\bigcup_{k,l} \left\{|\hat{P}_{k,l} - P_{k,l}| \geq \frac{\gamma}{\sqrt{n}}\right\} \mid N, \Gamma\right) \mathbb{P}(N) \mathbb{P}(\Gamma \mid N) + \mathbb{P}(\Gamma^c) + \mathbb{P}(N^c) \\ &= \mathbb{P}(\Gamma^c) + \mathbb{P}(N^c) \\ &\leq 2K^2 \exp\left(-\frac{nr^2}{2(B_1/4 + B_2r)}\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{n\pi_m^2}{8A_1\sigma^2 + 4A_2\pi_m}\right) \\ &\leq 2K^2 \exp\left(-\frac{\left(\frac{\pi_m^2\gamma}{5K} - \frac{1}{2}\right)^2}{2(B_1/4 + B_2\frac{\pi_m^2\gamma}{5K\sqrt{n}} - \frac{1}{2})}\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{n\pi_m^2}{8A_1\sigma^2 + 4A_2\pi_m}\right), \end{aligned}$$

where we apply Lemma 1.1 in the last inequality.

■

7.3 Proof of Theorem 3

Lemma 3 provides a more complete version of Theorem 3 by giving explicitly the constants.

Lemma 3. We keep the notations of Theorem 1.

Assume that $\alpha_n \log(n) \leq 1/L$ and that $n\alpha_n > \max\left(\frac{4Lc''}{\pi_m^2 D^2}, \frac{2}{L\pi_m}, \frac{2L\log(n)}{\pi_m c' D^2}\right)$. Then for all $t > 0$, it holds

$$\mathbb{P}(\|\hat{\pi} - \pi\|_\infty > t) \leq 2K \exp\left(-\frac{nt^2}{2(A_1/4 + A_2t)}\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{n\pi_m^2}{2A_1 + 4A_2\pi_m}\right).$$

Proof of Lemma 3.

Using Theorem 1.2 from [16], we get that

$$\forall c \in [K], \forall t > 0, \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=c} - \pi(c)\right| \geq t\right) \leq 2 \exp\left(-\frac{nt^2}{2(A_1\sigma_c^2 + A_2t)}\right)$$

where $A_1 = \frac{1 + (\lambda_+ \vee 0)}{1 - (\lambda_+ \vee 0)}$, $A_2 = \frac{5}{3} \mathbb{1}_{\lambda_+ \leq 0} + \frac{5}{1 - \lambda_+} \mathbb{1}_{\lambda_+ > 0}$ and $\sigma_c^2 = \pi(c)(1 - \pi(c)) \leq 1/4$.

We define the event $N := \left\{\text{err}(\hat{G}, G) < \exp\left(-\frac{\pi_m n \alpha_n c' D^2}{2L}\right)\right\}$. Note that on N , the partition of the clusters is correctly recovered thanks to the condition $n\alpha_n > \frac{2L \log(n)}{\pi_m c' D^2}$. Then,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k \in [K]} \{|\hat{\pi}(k) - \pi(k)| > t\}\right) &\leq \mathbb{P}\left(\bigcup_{k \in [K]} \{|\hat{\pi}(k) - \pi(k)| > t\} \mid N\right) + \mathbb{P}(N^c) \\ &= \mathbb{P}\left(\bigcup_{k \in [K]} \left\{\left|\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=k} - \pi(k)\right| > t\right\} \mid N\right) + \mathbb{P}(N^c) \\ &\leq 2K \exp\left(-\frac{nt^2}{2(A_1/4 + A_2t)}\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{n\pi_m^2}{2A_1 + 4A_2\pi_m}\right), \end{aligned}$$

where we apply Lemma 1.1 in the last inequality.

■

7.4 Proof of Theorem 4

We start by proving Lemma 4 which enriches the statement of Theorem 4 by giving explicitly the constants.

Lemma 4. We keep the notations of Theorem 1.

Assume that $\alpha_n \log(n) \leq \frac{1}{L}$ and that $n\alpha_n > \max\left(\frac{4Lc''}{\pi_m^2 D^2}, \frac{2}{L\pi_m}, \frac{2L \log(n)}{\pi_m c' D^2}\right)$. Then for all $0 < t < \pi_m - \frac{1}{n}$, it holds

$$\mathbb{P}(\|\hat{Q} - Q\|_\infty > t) \leq K(K+1) \exp\left(-\frac{(n\pi_m - nt - 1)^2 t^2}{\frac{1}{2} + \frac{2}{3}t}\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{nt^2}{2(A_1/4 + A_2t)}\right).$$

Proof of Lemma 4.

• Preliminary 1

Using Bernstein inequality, we get that

$$\forall k, l \in [K]^2 \text{ with } k \neq l, \forall t > 0, \mathbb{P}\left(\left|\frac{1}{|G_k| \times |G_l|} \sum_{i \in G_k} \sum_{j \in G_l} X_{i,j} - Q_{k,l}\right| \geq t\right) \leq 2 \exp\left(-\frac{|G_k| \times |G_l| t^2}{2(Q_{k,l}(1 - Q_{k,l}) + t/3)}\right)$$

and

$$\forall k \in [K], \forall t > 0, \mathbb{P}\left(\left|\frac{1}{|G_k| \times (|G_k| - 1)} \sum_{i, j \in G_k, i \neq j} X_{i,j} - Q_{k,k}\right| \geq t\right) \leq 2 \exp\left(-\frac{|G_k| \times (|G_k| - 1) t^2}{2(Q_{k,k}(1 - Q_{k,k}) + t/3)}\right).$$

- Preliminary 2

We define the event $N := \left\{ \text{err}(\hat{G}, G) < \exp\left(-\frac{\pi_m n \alpha_n c' D^2}{2L}\right) \right\}$. Note that on N , the partition of the clusters is correctly recovered thanks to the condition $n \alpha_n > \frac{2L \log(n)}{\pi_m c' D^2}$ and using Lemma 1.1.

- Preliminary 3

Using Theorem 1.2 from [16], we get that

$$\forall c \in [K], \forall t > 0, \quad \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=c} - \pi(c)\right| \geq t\right) \leq 2 \exp\left(-\frac{nt^2}{2(A_1/4 + A_2t)}\right).$$

We deduce that for all $t > 0$,

$$\mathbb{P}\left(\bigcup_c \left\{\left|\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=c} - \pi(c)\right| \geq t\right\}\right) \leq 2K \exp\left(-\frac{nt^2}{2(A_1/4 + A_2t)}\right).$$

We define $\Omega^c := \bigcup_{c \in [K]} \left\{\left|\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_i=c} - \pi(c)\right| \geq t\right\}$.

Considering $0 < t < \pi_m - \frac{1}{n}$, we have

$$\begin{aligned} \mathbb{P}(\|\hat{Q} - Q\|_\infty > t) &\leq \mathbb{P}\left(\bigcup_{k,l \in [K]^2, k \leq l} \{|\hat{Q}_{k,l} - Q_{k,l}| > t\} \mid \Omega\right) + \mathbb{P}(\Omega^c) \\ &\leq \mathbb{P}\left(\bigcup_{k,l \in [K]^2, k \leq l} \{|\hat{Q}_{k,l} - Q_{k,l}| > t\} \mid N, \Omega\right) + \mathbb{P}(N^c \mid \Omega) + \mathbb{P}(\Omega^c) \\ &\text{and using preliminary 3,} \\ &\leq \mathbb{P}\left(\bigcup_{k,l \in [K]^2, k \leq l} \{|\hat{Q}_{k,l} - Q_{k,l}| > t\} \mid N, \Omega\right) + \mathbb{P}(N^c \mid \Omega) + 2K \exp\left(-\frac{nt^2}{2(A_1/4 + A_2t)}\right) \\ &\leq \mathbb{P}\left(\bigcup_{k,l \in [K]^2, k \leq l} \{|\hat{Q}_{k,l} - Q_{k,l}| > t\} \mid N, \Omega\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{nt^2}{2(A_1/4 + A_2t)}\right) \\ &\text{where we used that } \mathbb{P}(N^c \mid \Omega) \leq \frac{c}{n^2} \text{ which is shown in the proof of Theorem 1,} \\ &\leq 2 \sum_{1 \leq k \leq l \leq K} \exp\left(-\frac{n(\pi(k) - t) \times (n\pi(l) - nt - 1)t^2}{2(Q_{k,l}(1 - Q_{k,l}) + t/3)}\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{nt^2}{2(A_1/4 + A_2t)}\right), \end{aligned}$$

where the last inequality is a direct consequence of the three preliminaries.

■

Proof of Theorem 4.

Let us consider $\gamma > 0$ and define $t = \frac{\gamma}{\sqrt{n}}$.

Asking

$$\frac{n \alpha_n}{\log(n)} \geq a \quad \text{with} \quad a := \frac{4Lc''}{c' \pi_m^2 D^2} \vee \frac{2L}{c' \pi_m D^2} \vee \frac{2}{L \pi_m},$$

we ensure that $n \alpha_n$ satisfies the conditions of Lemma 4.

Now let us look into the condition $t = \frac{\gamma}{\sqrt{n}} < \pi_m - \frac{1}{n}$ of Theorem 4. We will ask t to satisfy the stronger condition

$$t = \frac{\gamma}{\sqrt{n}} < \frac{\pi_m}{2} - \frac{1}{n} \quad \Leftrightarrow \quad 0 < \frac{\pi_m}{2} n - \gamma \sqrt{n} - 1.$$

Studying the polynomial function $f : x \mapsto \frac{\pi_m}{2} x^2 - \gamma x - 1$, one can find that the zeros of f are

$$x_1 := \frac{\gamma - \sqrt{\gamma^2 + 2\pi_m}}{\pi_m} \quad \text{and} \quad x_2 := \frac{\gamma + \sqrt{\gamma^2 + 2\pi_m}}{\pi_m} \leq \frac{2\gamma + \sqrt{2\pi_m}}{\pi_m}.$$

We deduce that asking

$$n > 4 \left(\frac{\gamma + 1}{\pi_m} \right)^2, \tag{11}$$

which implies $\sqrt{n} > \frac{2\gamma + \sqrt{2\pi_m}}{\pi_m}$, we guarantee that $\gamma/\sqrt{n} < \pi_m - 1/n$. Applying Lemma 4, we get that with probability at least

$$1 - \left(K(K+1) \exp\left(-\frac{(n\pi_m - \gamma\sqrt{n} - 1)^2 \frac{\gamma^2}{n}}{\frac{1}{2} + \frac{2}{3} \frac{\gamma}{\sqrt{n}}}\right) + \frac{c}{n^2} + 2K \exp\left(-\frac{\gamma^2}{2(A_1/4 + A_2 \frac{\gamma}{\sqrt{n}})}\right) \right),$$

it holds $\|\hat{Q} - Q\|_\infty \leq \gamma/\sqrt{n}$.

Thanks to (11), we have $(n\pi_m - \gamma\sqrt{n} - 1)^2 = n^2(\pi_m - \gamma/\sqrt{n} - 1/n)^2 \geq n^2\pi_m^2/4$ and $\gamma/\sqrt{n} \leq \pi_m/2$. We deduce that defining

$$b := c \vee (2K(K+1)) \quad \text{and} \quad b' := \frac{1}{2(A_1/4 + A_2\pi_m)} \wedge \frac{\pi_m^2}{2 + \frac{4}{3}\pi_m},$$

it holds with probability at least $1 - b(1/n^2 \vee \exp(-b'\gamma^2))$

$$\|\hat{Q} - Q\|_\infty \leq \frac{\gamma}{\sqrt{n}}.$$

■

Acknowledgments

This work was supported by grants from Région Ile-de-France. I would like also to thank my two PhD supervisors Claire Lacour and Yohann De Castro for their advice, their support and their time devoted to several proofreading and many useful discussions.

References

- [1] Emmanuel Abbe. *Community Detection and Stochastic Block Models*. Jan. 2018. ISBN: 9781680834772. DOI: [10.1561/9781680834772](https://doi.org/10.1561/9781680834772).
- [2] Emmanuel Abbe and Colin Sandon. *Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation*. 2016.
- [3] Emmanuel Abbe and Colin Sandon. “Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery”. In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE. 2015, pp. 670–688.
- [4] Alain Celisse, Jean-Jacques Daudin, and Laurent Pierre. “Consistency of maximum-likelihood and variational estimators in the stochastic block model”. In: *Electronic Journal of Statistics* 6.0 (2012), 1847–1899. ISSN: 1935-7524. DOI: [10.1214/12-ejs729](https://doi.org/10.1214/12-ejs729). URL: <http://dx.doi.org/10.1214/12-EJS729>.
- [5] Moses Charikar et al. “A Constant-Factor Approximation Algorithm for the k-Median Problem”. In: *Journal of Computer and System Sciences* 65 (Aug. 2002), pp. 129–149. DOI: [10.1006/jcss.2002.1882](https://doi.org/10.1006/jcss.2002.1882).
- [6] Yudong Chen and Jiaming Xu. *Statistical-Computational Tradeoffs in Planted Problems and Submatrix Localization with a Growing Number of Clusters and Submatrices*. 2014.
- [7] Sang Chin, Anup Rao, and Van Vu. *Stochastic Block Model and Community Detection in the Sparse Graphs: A spectral algorithm with optimal rate of recovery*. Jan. 2015.
- [8] Stéphane Chrétien, Clement Dombry, and Adrien Faivre. “A Semi-Definite Programming approach to low dimensional embedding for unsupervised clustering”. In: *Frontiers in Applied Mathematics and Statistics* 5 (June 2016). DOI: [10.3389/fams.2019.00041](https://doi.org/10.3389/fams.2019.00041).
- [9] Yingjie Fei and Yudong Chen. “Exponential error rates of SDP for block models: Beyond Grothendieck’s inequality”. In: *IEEE Transactions on Information Theory* PP (May 2017). DOI: [10.1109/TIT.2018.2839677](https://doi.org/10.1109/TIT.2018.2839677).
- [10] Elisabeth Gassiat, Alice Cleynen, and Stéphane Robin. “Inference in finite state space non parametric Hidden Markov Models and applications”. In: *Statistics and Computing* 26 (Jan. 2015). DOI: [10.1007/s11222-014-9523-8](https://doi.org/10.1007/s11222-014-9523-8).
- [11] Christophe Giraud and Nicolas Verzelen. “Partial recovery bounds for clustering with the relaxed K-means”. In: *Mathematical Statistics and Learning* 1.3 (May 2019), 317–374. ISSN: 2520-2316. DOI: [10.4171/msl/8](https://doi.org/10.4171/msl/8). URL: <http://dx.doi.org/10.4171/msl/8>.

-
- [12] Olivier Guédon and Roman Vershynin. “Community detection in sparse networks via Grothendieck’s inequality”. In: *Probability Theory and Related Fields* 165 (Nov. 2014). DOI: [10.1007/s00440-015-0659-z](https://doi.org/10.1007/s00440-015-0659-z).
 - [13] Bruce Hajek, Yihong Wu, and Jiaming Xu. *Semidefinite Programs for Exact Recovery of a Hidden Community*. Feb. 2016.
 - [14] Daniel J. Hsu, Sham M. Kakade, and Tong Zhang. “A Spectral Algorithm for Learning Hidden Markov Models”. In: *CoRR* abs/0811.4413 (2008). arXiv: [0811.4413](https://arxiv.org/abs/0811.4413). URL: <http://arxiv.org/abs/0811.4413>.
 - [15] Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi. “Phase Transitions in Semidefinite Relaxations”. In: *Proceedings of the National Academy of Sciences* 113 (Nov. 2015). DOI: [10.1073/pnas.1523097113](https://doi.org/10.1073/pnas.1523097113).
 - [16] Bai Jiang, Qiang Sun, and Jianqing Fan. “Bernstein’s inequality for general Markov chains”. In: (May 2018). arXiv: [1805.10721](https://arxiv.org/abs/1805.10721).
 - [17] Brian Karrer and M.E.J. Newman. “Stochastic blockmodels and community structure in networks”. In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 83 (Jan. 2011). DOI: [10.1103/PhysRevE.83.016107](https://doi.org/10.1103/PhysRevE.83.016107).
 - [18] Nicolas Keriven and Samuel Vaiter. *Sparse and Smooth: improved guarantees for Spectral Clustering in the Dynamic Stochastic Block Model*. 2020. arXiv: [2002.02892](https://arxiv.org/abs/2002.02892) [[stat.ML](#)].
 - [19] Jing Lei and Alessandro Rinaldo. “Consistency of spectral clustering in stochastic block models”. In: *The Annals of Statistics* 43.1 (Feb. 2015), 215–237. ISSN: 0090-5364. DOI: [10.1214/14-aos1274](https://doi.org/10.1214/14-aos1274). URL: <http://dx.doi.org/10.1214/14-AOS1274>.
 - [20] Catherine Matias and Vincent Miele. “Statistical clustering of temporal networks through a dynamic stochastic block model”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (June 2015). DOI: [10.1111/rssb.12200](https://doi.org/10.1111/rssb.12200).
 - [21] Jiming Peng and Yu Wei. “Approximating K-means-type Clustering via Semidefinite Programming”. In: *SIAM Journal on Optimization* 18 (Jan. 2007), pp. 186–205. DOI: [10.1137/050641983](https://doi.org/10.1137/050641983).
 - [22] Marianna Pensky and Teng Zhang. “Spectral clustering in the dynamic stochastic block model”. In: *Electronic Journal of Statistics* 13 (May 2017). DOI: [10.1214/19-EJS1533](https://doi.org/10.1214/19-EJS1533).
 - [23] William Perry and Alexander S. Wein. “A semidefinite program for unbalanced multisection in the stochastic block model”. In: *CoRR* abs/1507.05605 (2015). arXiv: [1507.05605](https://arxiv.org/abs/1507.05605). URL: <http://arxiv.org/abs/1507.05605>.
 - [24] Oleksandr Shchur and Stephan Günnemann. *Overlapping Community Detection with Graph Neural Networks*. 2019. arXiv: [1909.12201](https://arxiv.org/abs/1909.12201) [[cs.LG](#)].
 - [25] Kevin S. Xu. “Stochastic Block Transition Models for Dynamic Networks”. In: *CoRR* abs/1411.5404 (2014). arXiv: [1411.5404](https://arxiv.org/abs/1411.5404). URL: <http://arxiv.org/abs/1411.5404>.
 - [26] Tianbao Yang et al. “Detecting communities and their evolutions in dynamic social networks - A Bayesian approach”. In: *Machine Learning* 82 (Feb. 2011), pp. 157–189. DOI: [10.1007/s10994-010-5214-7](https://doi.org/10.1007/s10994-010-5214-7).

A Spectral gap for Markov chains

This section is largely inspired from [16] section 2.1.

We consider a state space E and a sigma-algebra Σ on E which is a standard Borel space. We denote by $(X_i)_{i \geq 1}$ a Markov chain on the state space (E, Σ) with invariant distribution π .

For any real-valued, Σ -measurable function $h : E \rightarrow \mathbb{R}$, we define $\pi(h) := \int h(x)\pi(dx)$. The set

$$\mathcal{L}_2(E, \Sigma, \pi) := \{h : \pi(h^2) < \infty\}$$

is a Hilbert space endowed with the inner product

$$\langle h_1, h_2 \rangle_\pi = \int h_1(x)h_2(x)\pi(dx), \quad \forall h_1, h_2 \in \mathcal{L}^2(E, \Sigma, \pi).$$

The map

$$\|\cdot\|_\pi : h \in \mathcal{L}_2(E, \Sigma, \pi) \mapsto \|h\|_\pi = \sqrt{\langle h, h \rangle_\pi},$$

is a norm on $\mathcal{L}_2(E, \Sigma, \pi)$. $\|\cdot\|_\pi$ naturally allows to define the norm of a linear operator T on $\mathcal{L}_2(E, \Sigma, \pi)$ as

$$N_\pi(T) = \sup\{\|Th\|_\pi : \|h\|_\pi = 1\}.$$

To each transition probability kernel $P(x, B)$ with $x \in E$ and $B \in \Sigma$ invariant with respect to π , we can associate a bounded linear operator $h \mapsto \int h(y)P(\cdot, dy)$ on $\mathcal{L}_2(E, \Sigma, \pi)$. Denoting this operator P , we get

$$Ph(x) = \int h(y)P(x, dy), \quad \forall x \in E, \quad \forall h \in \mathcal{L}_2(E, \Sigma, \pi).$$

Denoting by P^* the adjoint or time-reversal operator of the Markov operator P , we can define the self-adjoint operator $R = (P + P^*)/2$. Let $\mathcal{L}_2^0(\pi) := \{h \in \mathcal{L}_2(E, \Sigma, \pi) : \pi(h) = 0\}$. The spectrum of a self-adjoint Markov operator like R acting on $\mathcal{L}_2^0(\pi)$ is contained in $[-1, +1]$. The gap between 1 and the maximum of the spectrum of R is called the right \mathcal{L}_2 -spectral gap of P .

Definition 1. (Right \mathcal{L}_2 -spectral gap) A Markov operator P has right \mathcal{L}_2 -spectral gap $1 - \lambda_+(R)$ if the operator $R = (P + P^*)/2$ satisfies

$$\lambda_+(R) := \sup\{s : s \in \text{spectrum of } R \text{ acting on } \mathcal{L}_2^0(\pi)\} < 1.$$

B Partial recovery bound in SBM with fixed assignment of the communities

Verzelen and Giraud in [11] introduce a relaxed version of the K -means algorithms on the columns of the adjacency matrix. One specificity of their algorithm is the fact that they are working with the square of the adjacency matrix. This choice allows them to tackle problems outside of the assortative setting and with a wide set of possible connectivity matrices Q contrary to previous works.

Theorem B presents the result of Verzelen and Giraud in the SBM framework with a connectivity matrix $Q = \alpha_n Q_0$.

Theorem B. (see Theorem 2 in [11]).

Assume that $\|Q_0\|_\infty \leq L$. We define the signal-to-noise ratio $s^2 = \Delta^2/(\alpha_n L)$, where $\Delta^2 = \min_{k \neq j} \Delta_{k,j}^2$ with

$$\Delta_{k,j}^2 = \sum_l m_l (Q_{k,l} - Q_{j,l})^2 = \alpha_n^2 \sum_l m_l ((Q_0)_{k,l} - (Q_0)_{j,l})^2.$$

Then, there exist three positive constants c, c', c'' , such that for any $1/m \leq \alpha_n L \leq 1/\log(n)$,

$$\frac{1}{m} \leq \beta \leq \beta(\alpha_n L) := \frac{K^3}{n} e^{4n\alpha_n L}$$

and

$$s^2 \geq c'' n/m,$$

with probability at least $1 - c/n^2$,

$$\text{err}(\hat{G}, G) \leq e^{-c's^2}.$$

In particular, since

$$s^2 = \frac{\alpha_n \min_{k \neq j} \sum_{l \in [K]} m_l ((Q_0)_{k,l} - (Q_0)_{j,l})^2}{L} \geq \frac{\alpha_n m D^2}{L},$$

we get that with probability at least $1 - c/n^2$,

$$-\log(\text{err}(\hat{G}, G)) = \Omega(m\alpha_n).$$