Sparse and smooth: Improved guarantees for spectral clustering in the dynamic stochastic block model[*]

Nicolas Keriven[1] and Samuel Vaiter[2]

[1]*CNRS & GIPSA*

[2]*CNRS & LJAD, Université Côte d'Azur*

Abstract: In this paper, we analyze classical variants of the Spectral Clustering (SC) algorithm in the Dynamic Stochastic Block Model (DSBM). Existing results show that, in the relatively sparse case where the expected degree grows logarithmically with the number of nodes, guarantees in the static case can be extended to the dynamic case and yield improved error bounds when the DSBM is sufficiently *smooth* in time, that is, the communities do not change too much between two time steps. We improve over these results by drawing a new link between the sparsity and the smoothness of the DSBM: the smoother the DSBM is, the more sparse it can be, while still guaranteeing consistent recovery. In particular, a mild condition on the smoothness allows to treat the *sparse* case with bounded degree. These guarantees are valid for the SC applied to the adjacency matrix or the normalized Laplacian. As a by-product of our analysis, we obtain to our knowledge the best spectral concentration bound available for the normalized Laplacian of matrices with independent Bernoulli entries.

MSC 2010 subject classifications: Primary 62F12; secondary 05C80, 62H30.
Keywords and phrases: Dynamic network, dynamic Stochastic Block Model, spectral Clustering, concentration bounds.

Received November 2020.

1. Introduction

In recent years, the study of dynamic networks has appeared as a topic of great interest to model complex phenomenons that evolve with time, such as interactions in social networks, the spread of infectious diseases or opinions, or information packets in computer networks. In light of this, many random graphs models, traditionally static (non-dynamic), have been extended to the dynamic case, see [16, 20] for reviews. One of the most popular use of dynamic networks consists in detecting and tracking communities of well-connected nodes, for instance users of a social network [49, 47, 44], also known as clustering. In this context, the classical Stochastic Block Model (SBM) [19], in which nodes intra- and inter-communities are linked independently with some prescribed probabilities,

has been extended to dynamic settings under the name of Dynamic Stochastic Block Model (DSBM) in a myriad of ways. In this paper, we consider one of the first, and most popular, extension [49] as a discrete Hidden Markov Model (HMM) as well as one of its simplification, where node memberships follow a Markov chain with respect to time, and connections are generated by a classical SBM conditionally on the memberships. We will also consider a slight simplification as in [37], where the authors remove the Markov Chain assumption and consider deterministic community memberships at each time steps. Many other models have been proposed since, to take into account evolving connection probabilities [47, 30, 37], varying number of nodes [46], connections that depend on their previous states [46], mixed-membership SBM [18] or multi-graphs [17].

The literature on clustering nodes in a graph is vast, with a variety of methods. Arguably, the most popular class of algorithms in practice is that of *Spectral Clustering* (SC) methods [32, 42], which consist in applying a classical clustering algorithm for vectorial data, often the well-known k-means algorithm [28], to the eigenvectors of a matrix related to the structure of the graph such as the adjacency matrix or (normalized) Laplacian. In a dynamic context, we will consider in this paper one of the simplest adaptation of SC, which consists in feeding a version of the adjacency matrix *smoothed in time* to the classical SC algorithm, in hope of implicitly enforcing smoothness of the communities. This can be an averaged version of the adjacency matrix over a finite window [18, 37], or computed through recursive updates with a certain "forgetting factor" [8, 9, 45], which is somehow more amenable to streaming computing. Other works explicitly enforce smoothness between the communities or between the eigenvectors considered in SC through efficient updates [33, 12, 27].

Beyond SC, many other methods have been proposed, such as Maximum Likelihood or variational approaches, which are consistent for the SBM and DSBM [6, 30, 29], Bayesian approaches [49], learning-based approaches [2], or neural networks [5]. Many variants of the SC itself exist, often to accelerate computation [41]. We focus here on the traditional SC.

Guarantees for Spectral Clustering There is a vast literature on the theoretical analysis of SC, and guarantees come in many different flavors. Several works analyze the algorithm when the graph is, in some sense, well-clustered [36], in terms of spectral convergence of the normalized Laplacian when the number of nodes goes to infinity [43, 13, 14, 40], or using random matrix theory [11].

It is well-known that a key quantity to analyze SC algorithms is the density of edges with respect to the number of nodes. In the specific case of independent Bernoulli edges like the SBM and DSBM, this correspond to the mean probability of connection, which will be denoted by α_n in this paper, where n is the number of nodes in the graph. The *dense* case $\alpha_n \sim 1$ is generally simple to analyze [24]. At the other end of the spectrum, the so-called *sparse* case $\alpha_n \sim \frac{1}{n}$ is much more complex, since the graph is not even guaranteed to be connected with high probability [1].

Modern analyses of the sparse case are often inspired by statistical physics [21, 31, 1], and are interested with the computation of a detectability threshold, that

is, the characterization of regimes of parameters in which there exists (or not) an algorithm that asymptotically performs better than random guess. However, this approach does not concern the classic SC algorithm (which will generally fail [21]), and the case where the number of communities K is larger than 2 is still largely open. In the dynamic case, a conjecture on the detectability threshold is given in [15]. In parallel, other works study the sparse case by regularizing the adjacency matrix or normalized Laplacian of the graph before the SC algorithm [23, 24, 7].

In [25], Lei and Rinaldo provide strong, non-asymptotic consistency guarantees for the classic SC algorithm on the adjacency matrix (without regularization) in the *relatively sparse* case $\alpha_n \gtrsim \frac{\log n}{n}$, showing that the proportion of misclassified nodes tends to 0 with a probability that goes to 1 when the number of nodes n increases. Their recovery results are valid for any K, potentially growing slowly with n. In [37], Pensky and Zhang extend this analysis to a particular DSBM, referred to as "deterministic" DSBM in the sequel, for the SC algorithm applied to the adjacency matrix smoothed over a finite window. In this case, another key quantity is the temporal *smoothness* of the model ε_n, that is, the proportion of nodes that may change community between two time steps (the smaller ε_n is, the smoother the model). They showed that, in the relatively sparse case, if the model was sufficiently smooth in $\varepsilon_n = o\left(\frac{1}{\log n}\right)$, then the error bound of the static case can be improved. However their analysis still takes place in the relatively sparse case even when ε_n is very low. In [4, 35], the authors consider constant cluster membership $\varepsilon_n = 0$, but changing probabilities of connection.

Contributions   In this paper, we follow the analyses of [25] and [37] and significantly extend them in several ways.

- Our main contribution is to draw a new link between the sparsity α_n and smoothness ε_n in the analysis of the DSBM: we show that, *the smoother the model, the sparser it can be*, while still guaranteeing consistency. In particular, a mildly strengthened condition $\varepsilon_n \sim \frac{1}{\log^2 n}$ allows to *give consistent guarantees in the sparse case $\alpha_n \sim \frac{1}{n}$*.
- We extend the result to the normalized Laplacian. As a by-product, in the static case, we obtain, to our knowledge, the best spectral concentration bound available (of order $\frac{1}{\sqrt{\log n}}$) in the relatively sparse case $\alpha_n \sim \frac{\log n}{n}$.
- We also improve the rate of the error bounds with respect to the number of communities K when the probabilities of connection between communities decrease with K, in both the static [25] and dynamic [37] cases.
- Finally, we extend our results to the Markov DSBM introduced in [49], and the SC algorithm with an "exponentially smoothed" matrix, used in [8, 45] and appropriate in a streaming computing framework.

Outline   In Section 2, we introduce notations, the SBM and DSBM, and recall the SC algorithm. In Section 3, we draw a link between recovery guarantees for SC and the concentration of the input matrix in spectral norm, similar to [25]

but extended to the normalized Laplacian. In Section 4 and 5, we expose our main concentration results respectively for the adjacency matrix and normalized Laplacian. Proofs are given in Section 6, with technical computations deferred to the Appendix. We provide some outlooks in Section 7.

2. Framework and notations

The set of the first n integers is denoted by $[n] = \{1, \ldots, n\}$. For any vector $d \in \mathbb{R}^n$, we define $\mathrm{diag}(d) \in \mathbb{R}^{n \times n}$ to be the diagonal matrix whose elements are given by d. For a varying parameter α_n, the notation $\alpha_n \sim f(n)$ indicates that, as $n \to \infty$, the quantity $\alpha_n/f(n)$ tends to a non-zero constant, $\alpha_n \lesssim f(n)$ indicates that there is a universal constant C such that $\alpha_n \leqslant Cf(n)$, and similarly for $\alpha_n \gtrsim f(n)$.

An undirected graph $G = (V, E)$ is formed by a set of nodes V and edges $E \subset V \times V$. For a graph with n nodes, we often adopt $V = [n]$, and we define its (symmetric) adjacency matrix $A \in \{0, 1\}^{n \times n}$ such that for $i, j \in [n]$,

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

We also define the (diagonal) degree matrix $D(A)$ by

$$D(A) = \mathrm{diag}\left((d_i)_{i=1}^n\right) \quad \text{where} \quad d_i = \sum_{j=1}^n A_{ij}.$$

For any symmetric matrix A such that $\sum_j A_{ij} \neq 0$ for all i, the normalized Laplacian $L(A)$ is defined as

$$L(A) = D(A)^{-\frac{1}{2}} A D(A)^{-\frac{1}{2}}.$$

We note that, typically in the litterature [10], the normalized Laplacian is defined as the symmetric matrix $\mathrm{Id} - D(A)^{-\frac{1}{2}} A D(A)^{-\frac{1}{2}}$. However, SC is mainly concerned with the eigenvectors of the Laplacian, which are the same for both variants.

Stochastic Block Model Let us start by introducing the classical static SBM. We take the following notations: n the number of nodes, K the number of communities. Each node belongs to exactly one community. We denote by $\Theta \in \mathcal{M}_{n,K} \subset \{0, 1\}^{n \times K}$ the $0-1$ matrix representing the memberships of nodes, where for each node i, $\Theta_{ik} = 1$ indicates that it belongs to the kth community, and is 0 otherwise, and $\mathcal{M}_{n,K}$ is the set of all such membership matrices. Given Θ, for $i < j$, we have

$$A_{ij} \mid \{\Theta_{ik} = 1, \Theta_{j\ell} = 1\} \sim \mathrm{Ber}(B_{k\ell}),$$

where $B \in [0, 1]^{K \times K}$ is a symmetric connectivity matrix, and $\mathrm{Ber}(p)$ indicates a Bernoulli random variable with parameter p. Finally, we define $P = \Theta B \Theta^\top \in$

$\mathbb{R}^{n \times n}$ the matrix storing the probabilities of connection between two nodes off its diagonal, and we have

$$\mathbb{E}(A) = P - \operatorname{diag}(P).$$

Typically, B has high diagonal terms and low off-diagonal terms. We will consider B of the form

$$B = \alpha_n B_0, \tag{1}$$

for some $\alpha_n \in (0, 1)$ and $B_0 \in [0, 1]^{K \times K}$ whose elements are denoted by $b_{k\ell}^{(0)}$. It is known that the rate α_n when $n \to \infty$ is the main key quantity when analyzing the properties of random graphs. Typical settings include $\alpha_n \sim 1$ (dense graphs), $\alpha_n \sim 1/n$ (sparse graphs), or middle grounds such as $\alpha_n \sim \frac{\log n}{n}$, usually referred to "relatively sparse" graphs. As we will see, it is known that strong guarantees of consistency can be given in the relatively sparse case, while the sparse case it hard to analyze and only partially understood.

For some maximum and minimum community sizes $n_{\max} \geqslant \frac{n}{K}$ and $n_{\min} \leqslant \frac{n}{K}$, we define the set of admissible community sizes $N \stackrel{\text{def.}}{=} \{(n_k)_{k=1}^K \mid n_{\min} \leqslant n_k \leqslant n_{\max}, \sum_k n_k = n\}$, and

$$\bar{n}_{\max} \stackrel{\text{def.}}{=} \max_{(n_\ell)_\ell \in N, k \leqslant K} \sum_\ell n_\ell b_{k\ell}^{(0)}, \quad \bar{n}_{\min} \stackrel{\text{def.}}{=} \min_{(n_\ell)_\ell \in N, k \leqslant K} \sum_\ell n_\ell b_{k\ell}^{(0)}. \tag{2}$$

These quantities are such that the expected degree will be comprised between $\alpha_n \bar{n}_{\min}$ and $\alpha_n \bar{n}_{\max}$. For simplicity, we will sometimes express our results with B_0 equal to:

$$B_0 = (1 - \tau)\operatorname{Id}_K + \tau 1_K 1_K^\top. \tag{3}$$

In other words, B contains α_n on its diagonal and $\tau \alpha_n$ outside. For this expression of B_0, we have $\bar{n}_{\max} = (1-\tau)n_{\max} + n\tau$, and similarly for $\bar{n}_{\min}$. Interestingly, in the case of balanced communities $n_{\max}, n_{\min} \sim \frac{n}{K}$, we have then

$$\bar{n}_{\min}, \bar{n}_{\max} \sim \begin{cases} n & \text{if } \tau \sim 1, \\ \frac{n}{K} & \text{if } \tau \sim \frac{1}{K}. \end{cases}$$

Dynamic SBM   The Dynamic SBM (DSBM) is a random model for generating adjacency matrices $A_0, \ldots, A_t$ at each time step. Each A_i will be generated according to a classical SBM with constant number of nodes n, number of communities K and connectivity matrix B, but changing node memberships Θ_t. Note that several works consider changing number of nodes [46] or changing connectivity matrix [37], but for simplicity we assume that they are constant in time here. Note that, using simple triangular inequalities as in [37], it would not be difficult to integrate a changing connectivity matrix B into our results. We will consider two potential models on the Θ_t.

– The simplest one, adopted in [37], is to consider that $\Theta_0, \ldots, \Theta_t$ are deterministic variables. In this case, we will assume that only a number $s \leqslant n$ of nodes change communities between each time step $t-1$ and t, and denote

$\varepsilon_n = s/n$ this relative proportion of nodes. We will also assume that at all time steps, the communities sizes are comprised between some $n_{\min}$ and $n_{\max}$, which will typically be of the order of n/K for balanced communities. As a shorthand, we will simply refer to this model as *deterministic* DSBM (keeping in mind that the A_t are still random).

– In the second model, similar to [49] we assume that the nodes memberships follow a Markov chain, such that between two time steps, all nodes have a probability $1 - \varepsilon_n$ to stay in the same community, and ε_n to go into any other community, that is:

$$\forall i, k, \ \mathbb{P}\left((\Theta_t)_{ik} = 1 \mid (\Theta_{t-1})_{ik} = 1\right) = 1 - \varepsilon_n,$$

$$\forall \ell \neq k, \ \mathbb{P}\left((\Theta_t)_{i\ell} = 1 \mid (\Theta_{t-1})_{ik} = 1\right) = \frac{\varepsilon_n}{K - 1}.$$

Then, conditionally on Θ_t, the A_t are drawn independently according to a SBM. The global model is thus a Hidden Markov Model (HMM). We will simply refer to this case as Markov DSBM. Note that, in this case, it is rather difficult to quantify, in a non-asymptotic manner, the probability of having bounded community sizes globally holding for all time steps. Hence $\bar{n}_{\max}, \bar{n}_{\min}$ will not intervene in our analysis of this case.

Goal and error measure The goal of a clustering algorithm is to give an estimator $\hat{\Theta}$ of the node memberships Θ, up to permutation of the communities labels. We consider the following measure of discrepancy between Θ and an estimator $\hat{\Theta}$ [25]:

$$E(\hat{\Theta}, \Theta) = \min_{Q \in \mathcal{P}_k} \frac{1}{n} \|\hat{\Theta}Q - \Theta\|_0, \tag{4}$$

where $\mathcal{P}_k$ is the set of permutation matrices of $[k]$ and $\|\cdot\|_0$ counts the number of non-zero elements of a matrix. While other error measures are possible, as we will see one can generally relate them to a spectral concentration property, which will be the main focus of this paper.

In the dynamic case, a possible goal is to estimate $\Theta_1, \ldots, \Theta_t$ for all time steps simultaneously [45, 37]. Here we consider a slightly different goal: *at a given time step t*, we seek to estimate Θ_t with the best precision possible, by exploiting past data. In general, this will give rise to methods that are computationally lighter than simultaneous estimation of all the Θ_t's, and more amenable to streaming computing, where one maintains an estimator without having to keep all past data in memory. Naturally, such methods could be applied independently at each time step to produce estimators of all the Θ_t's, but this is not the primary goal here.

Spectral Clustering (SC) algorithm Spectral Clustering [32] is nowadays one of the leading methods to identify communities in an unsupervised setting. The basic idea is to solve the K-means problem [28] on the K leading eigenvectors E_K of either the adjacency matrix or (normalized) Laplacian. Solving the

---

Algorithm 1: Spectral Clustering algorithm

Data: Matrix $M \in \mathbb{R}^{n \times n}$ (typically adjacency or normalized Laplacian), number of
communities K, approximation ratio $\delta > 0$.
Result: Estimated communities $\hat{\Theta} \in \mathbb{R}^{n \times K}$.

1 Compute the K leading eigenvectors E_K of M.
2 Obtain a $(1 + \delta)$-approximation $(\hat{\Theta}, \hat{C})$ of (5).
3 Return $\hat{\Theta}$.

K-means, i.e., obtaining

$$(\bar{\Theta}, \bar{C}) \in \underset{\Theta \in \mathcal{M}_{n,K}, C \in \mathbb{R}^{K \times K}}{\operatorname{Argmin}} \|\Theta C - E_K\|_F^2, \tag{5}$$

is known to be NP-hard, but several approximation algorithms, such as [22], are
known to produce $1 + \delta$ approximate solutions $(\hat{\Theta}, \hat{C})$

$$\|\hat{\Theta}\hat{C} - E_K\|_F^2 \leqslant (1 + \delta)\|\bar{\Theta}\bar{C} - E_K\|_F^2.$$

The SC is summarized in Algorithm 1.

In the dynamic case, a typical approach to exploit past data is to replace the
adjacency matrix A_t with a version "smoothed" in time A_t^{smooth}, and feed either
$\hat{P} = A_t^{\mathrm{smooth}}$ or the corresponding Laplacian $\hat{L} = L(A_t^{\mathrm{smooth}})$ to the classical
SC algorithm. In [37], the authors consider the smoothed adjacency matrix as
an average over its last r values:

$$A_t^{\mathrm{unif}} = \frac{1}{r} \sum_{k=0}^{r-1} A_{t-k}. \tag{6}$$

Note that, in the original paper, the authors sometimes consider non-uniform
weights due to potential changes in time of the connectivity matrix B_t, but in
our case we consider a fixed B, and thus uniform weights $\frac{1}{r}$. In this paper, we
will also consider the "exponentially smoothed" estimator proposed by [8, 9, 48],
which is computed recursively as:

$$A_t^{\mathrm{exp}} = (1 - \lambda)A_{t-1}^{\mathrm{exp}} + \lambda A_t. \tag{7}$$

for some "forgetting factor" $\lambda \in (0, 1]$, and $A_0^{\mathrm{exp}} = A_0$. Compared to the uniform
estimator (6), this kind of estimator is somewhat more amenable to streaming
and online computing, since only the current A_t^{exp} needs to be stored in memory
instead of the last r values $A_t, A_{t-1}, \ldots, A_{t-r+1}$. Note however that A_t^{exp} may
be denser than a typical adjacency matrix, so the memory gain is sometimes
mitigated depending on the case. Note that in fact A^{unif} and A^{exp} can be both
written as a weighted sum $A_t^{\mathrm{smooth}} = \sum_{k=0}^{t} \beta_k A_{t-k}$, and our results are actually
valid for any weights β_k satisfying some assumptions, see eq. (25). Uniform and
exponential weights are most commonly used in practice.
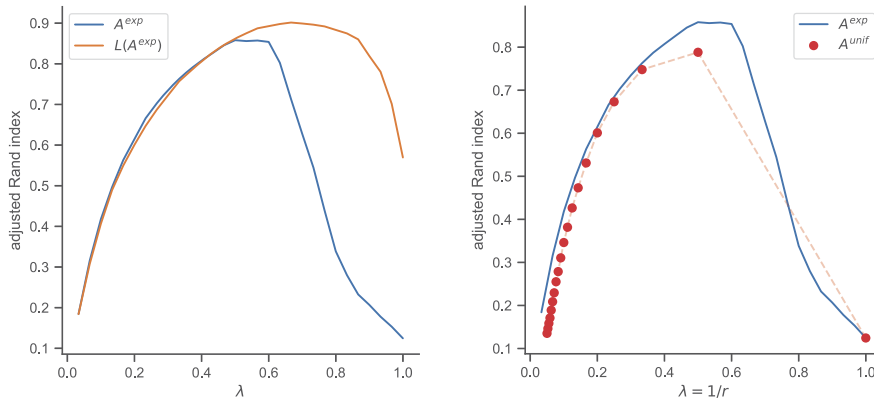
(a) Adjusted Rand index in function of the forgetting factor λ for the adjacency matrix and the normalized Laplacian.

(b) Adjusted Rand index in function of the forgetting factor λ for $A_t^{\exp}$ and the window size r for A_t^{unif}.

FIG 1. *Performance results for SC on synthetic data.*

In Fig. 1, we illustrate the performance of the SC algorithm on a synthetic DSBM example. As expected [39], the normalized Laplacian $L(A_t^{\exp})$ generally performs better than $A_t^{\exp}$. Interestingly, the optimal forgetting factor λ is slightly different from one to the other, and the normalized Laplacian reaches a higher performance altogether. We then compare A_t^{unif} and $A_t^{\exp}$. As we will see in the sequel, taking $r \sim \frac{1}{\lambda}$ often results in the same performance for both estimators. However, a clear advantage of the exponential estimator is that it is not limited to discrete window sizes, but has a continuous forgetting factor. As such, $A_t^{\exp}$ with the optimal λ often reaches a better performance than A_t^{unif} with the optimal r.

3. From Spectral Clustering to spectral norm concentration

As described in [25], a key quantity for analyzing SC algorithm is the concentration of the adjacency matrix around its expectation *in spectral norm*. As a first contribution, we prove the following lemma, which is a generalisation of this result to the normalized Laplacian.

Lemma 1. *Let $P = \Theta B \Theta^\top$ correspond to some SBM with K communities, where $n_{\max}$, $n'_{\max}$ and $n_{\min}$ are respectively the largest, second-largest and smallest community size. Assume $B = \alpha_n B_0$ for any B_0 with smallest eigenvalue γ. Let $\hat{P}$ be an estimator of P, and $\hat{\Theta}$ be the output of Algorithm 1 on $\hat{P}$ with a $(1 + \delta)$-approximate k-means algorithm. Then*

$$E(\hat{\Theta}, \Theta) \lesssim (1 + \delta) \frac{n'_{\max} K}{n \alpha_n^2 n_{\min}^2 \gamma^2} \| \hat{P} - P \|^2 \,, \tag{8}$$

Similarly, if $\hat{L}$ is an estimator of $L(P)$ and $\hat{\Theta}$ is the output of Algorithm 1 on $\hat{L}$, it holds that

$$E(\hat{\Theta}, \Theta) \lesssim (1 + \delta) \frac{n'_{\max} K \bar{n}_{\max}^2}{n n_{\min}^2 \gamma^2} \|\hat{L} - L(P)\|^2 . \tag{9}$$

When B_0 is defined as (3), we have $\gamma = 1 - \tau$.

The proof of this lemma is deferred to Appendix A.1. The first bound (8) was proved in [25], we extend it to the Laplacian case. Note that $\hat{L}$ could be an estimator of $L(P)$ without being of the form $\hat{L} = L(M)$ for some matrix M.

Using this lemma, in the static SBM case, the goal is to find estimators $\hat{P}$ or $\hat{L}$ that concentrate around P or $L(P)$ in spectral norm. In the dynamic case, where the goal is to estimate the communities at a particular time t, we seek the best estimators for P_t or $L(P_t)$. As outlined in the previous section, we will consider smoothed versions of the adjacency matrix A_t^{smooth}, and prove concentration of A_t^{smooth} around P_t and $L(A_t^{\text{smooth}})$ around $L(P_t)$.

Remark 1. Assuming that all community sizes are of the order of $\frac{n}{K}$ and τ is fixed, the error in the adjacency case (8) scales as $\frac{K^2}{n^2 \alpha_n^2} \|\hat{P} - P\|^2$, and in the normalized Laplacian case the error (9) scales as $K^2 \|\hat{L} - L(P)\|^2$. Also note that, when $\bar{n}_{\max} \sim \frac{n}{K}$, then the error (9) is as $\|\hat{L} - L(P)\|^2$. This does not explicitly depend on α_n or K, however these quantities will naturally appear in the concentration of the Laplacian.

The next sections will therefore be devoted in analyzing the spectral concentration rates of the various estimators. Table 1 summarize our results and compare them with previous works. As we will see in the next section, our main contribution is to weaken the hypothesis on the sparsity α_n, and relate it to the smoothness of the DSBM ε_n. We also provide the best bound available for the normalized Laplacian in the static case, and the first bound in the dynamic case.

In Figure 2, we illustrate numerically the spectral concentration of A_t^{exp} and $L(A_t^{\text{exp}})$, and their actual clustering performance, with respect to the forgetting factor λ. We see that there is a slight discrepancy between the λ that minimizes the spectral bound, and the one that yields the best clustering result. As we will see in the next sections, the λ that minimizes the spectral error is theoretically of the order of $\sqrt{\alpha_n n \varepsilon_n}$. This rate is indeed verified numerically for the spectral error, however the actual best clustering performance deviates slightly. This indicates that spectral norm concentration probably does not yield sharp bounds in examining the performance of k-means and SC. We consider this to be an open, difficult question, as many analyses of k-means rely on spectral properties. This is left for future investigations.

4. Spectral concentration of the adjacency matrix

We start by recalling the result of [25] in the static case and prove an interesting improvement in some cases, then we examine the result for DSBM of [37] and
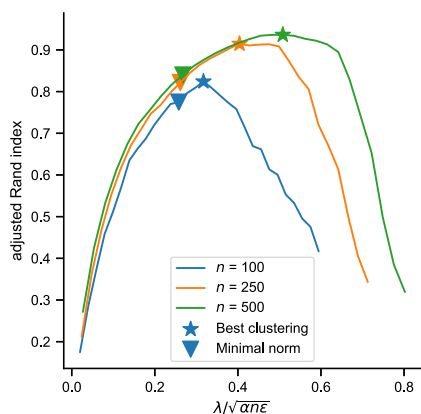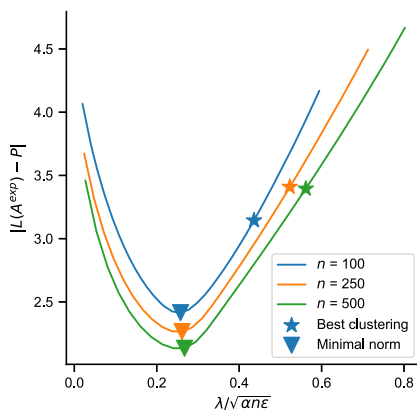
(a) Adjusted Rand index in function of the normalized forgetting factor $\lambda/\sqrt{\alpha_n n \varepsilon_n}$ for the adjacency matrix.
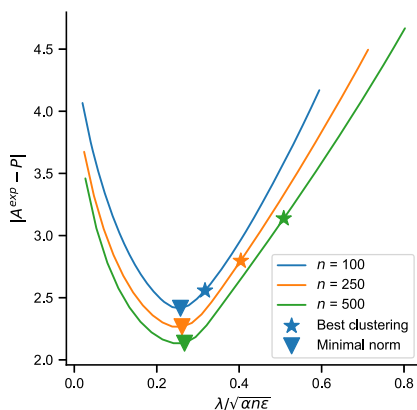
(b) Approximation (in norm) of P by $A^{\exp}$ in function of the normalized forgetting factor $\lambda/\sqrt{\alpha_n n \varepsilon_n}$ for the adjacency matrix.
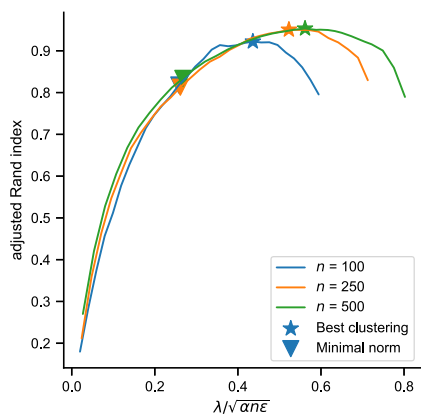
(c) Adjusted Rand index in function of the normalized forgetting factor $\lambda/\sqrt{\alpha_n n \varepsilon_n}$ for the normalized Laplacian matrix.

(d) Approximation (in norm) of P by $L(A^{\exp})$ in function of the normalized forgetting factor $\lambda/\sqrt{\alpha_n n \varepsilon_n}$ for the normalized Laplacian matrix.

Fig 2. *SC on synthetic data. Comparison between the forgetting factor λ that minimizes the spectral error, and the one that yields the best clustering result.*

	Static		Dynamic			
	Adjacency	Laplacian	Adjacency	Laplacian	Hypothesis	$E \to 0$
[34]	$\log n$	$\mathcal{O}(1)$			$\alpha_n \gtrsim \frac{\log n}{n}$	Yes
[25]	$\sqrt{\alpha_n n}$				$\alpha_n \gtrsim \frac{\log n}{n}$	Yes
[37]	$\sqrt{\alpha_n n}$		$\sqrt{\alpha_n n \rho_n}$		$\alpha_n \gtrsim \frac{\log n}{n}$	Yes
[3]	$\sqrt{\log n}$				$\alpha_n \gtrsim \frac{1}{n}$	No
Us	$\sqrt{\alpha_n n}$	$\frac{1}{\sqrt{\alpha_n n}}$	$\sqrt{\alpha_n n \rho_n}$	$\sqrt{\frac{\rho_n}{\alpha_n n}}$	$\frac{\alpha_n}{\rho_n} \gtrsim \frac{\log n}{n}$	Yes

state our main contribution, that is, a weakening of the sparsity hypothesis for
this case.

4.1. Static case

In their landmark paper [25], Lei and Rinaldo analyze the relatively sparse case
$\alpha_n \gtrsim \frac{\log n}{n}$ and show that, with probability at least $1 - n^{-\nu}$ for some $\nu > 0$, the
adjacency matrix concentrates as

$$\|A - P\| \lesssim \sqrt{n\alpha_n}. \tag{10}$$

Therefore, by Lemma 1, using A as an estimator for P in an SC algorithm leads
to an error $E(\hat{\Theta}, \Theta) \lesssim \frac{K^2}{\alpha_n n}$, such that $E(\hat{\Theta}, \Theta) \to 0$ whenever $K = o(\sqrt{n\alpha_n})$.
As a minor contribution, we remark that it is not hard to prove the following
Proposition that improves over their result in the regime where α_n is slightly
larger than $\frac{\log n}{n}$ and B_0 is defined as (3).

Proposition 1. *Consider a static SBM where B_0 is defined as* (3)*, assume that
the community sizes $n_1, \ldots, n_K$ are comprised between $n_{\min}$ and $n_{\max}$, and that*

$$\alpha_n \gtrsim \frac{\log n}{\bar{n}_{\min}}. \tag{11}$$

*Then, for all $\nu > 0$, there exists a constant C_ν such that, with probability at
least $1 - \sum_k n_k^{-\nu}$, it holds that*

$$\|A - P\| \leqslant C_\nu \sqrt{\bar{n}_{\max}\alpha_n}. \tag{12}$$

Proof. Denote by $S_1, \ldots, S_K \subset [n]$ the subset of indices of each community,
assume without lost of generality that the nodes are ordered such that the S_k
are consecutive in $[n]$, that is, $S_1 = \{1, \ldots, n_1\}$, $S_2 = \{n_1 + 1, \ldots, n_1 + n_2\}$, and

so on. Define $A_k = A_{S_k,S_k} \in \{0,1\}^{n_k \times n_k}$ the adjacency matrix of the subgraph of nodes from the kth community. Note that by our assumption on B we have $P_k = P_{S_k,S_k} = \alpha_n 1_{n_k} 1_{n_k}^\top$. Denote $A' \in \{0,1\}^{n \times n}$ the block matrix containing the A_k on its diagonal of blocks, similarly P', and $A'' = A - A'$, $P'' = P - P'$. We have

$$\|A - P\| \leqslant \|A' - P'\| + \|A'' - P''\| = \max_k \|A_k - P_k\| + \|A'' - P''\|,$$

where the equality is valid because $A' - P'$ is a block diagonal matrix. From Lei and Rinaldo's result above, for each k, if $\alpha_n \gtrsim \frac{\log n_k}{n_k}$, then with probability at least $1 - n_k^{-\nu}$ it holds that $\|A_k - P_k\| \lesssim \sqrt{n_k \alpha_n}$, such that $\|A' - P'\| \lesssim \sqrt{n_{\max} \alpha_n}$. For the second term, we note that A'' is an adjacency matrix generated by the SBM corresponding to P'', whose maximal probability is $\tau \alpha_n$. Hence, if $\tau \alpha_n \gtrsim \frac{\log n}{n}$, then with probability $1 - n^{-\nu}$ we have $\|A'' - P''\| \lesssim \sqrt{\tau n \alpha_n}$. We conclude with a union bound. □

This proposition provides a better error rate than [25] when τ goes to 0 with K, at the price of requiring a higher α_n. For instance, when the communities sizes are balanced $n_k \sim \frac{n}{K}$, and we have $\tau \sim \frac{1}{K}$ and $\alpha_n \sim \frac{K \log n}{n}$, Lei and Rinaldo's rate (10) yields $E(\hat{\Theta}, \Theta) \lesssim \frac{K}{\log n}$ and converge only for $K = o(\log n)$, while using Proposition 1 we get $E(\hat{\Theta}, \Theta) \lesssim \frac{1}{\log n}$. The latter does not depend on K, which brings a strict improvement compared to (10). Recall however that τ and α_n do depend on K, and that there must be a $\nu > 0$ such that $K^{\nu+1} n^{-\nu} \to 0$ to obtain a probability rate that goes to 1.

4.2. Dynamic case

In [37], Pensky and Zhang analyze the dynamic case with Lei and Rinaldo's proof technique. They consider the deterministic DSBM model in the almost sparse case $\alpha_n \gtrsim \frac{\log n}{n}$ and the uniform estimator (6). Defining a factor

$$\rho_n^{(\mathrm{PZ})} = \min(1, \sqrt{n \alpha_n \varepsilon_n}), \tag{13}$$

they show that, for an optimal choice of window size $r \sim \frac{1}{\rho_n^{(\mathrm{PZ})}}$, it holds that

$$\|A_t^{\mathrm{unif}} - P_t\| \lesssim \sqrt{n \alpha_n \rho_n^{(\mathrm{PZ})}}. \tag{14}$$

which is valid for t sufficiently large to avoid degenerate situations, in conditions similar to that of Theorem 1 below. In particular, the concentration is better if $\rho_n^{(\mathrm{PZ})} = o(1)$, that is:

$$\varepsilon_n = o\left(\frac{1}{\alpha_n n}\right). \tag{15}$$

In other words, there is an improvement if we assume sufficient *smoothness* in time, which then leads to a better error rate $E(\hat{\Theta}, \Theta) \lesssim \frac{K^2 \rho_n^{(\mathrm{PZ})}}{\alpha_n n}$ when using

A_t^{unif} in the SC algorithm. Note that, with this proof technique, a constant smoothness $\varepsilon_n \sim 1$ does not improve the error rate (see remark 2).

We remark that, despite the assumption on the smoothness and the availability of more data, the result above still assumes the relative sparse case. However, with sufficient smoothness, it should be possible to weaken the hypothesis made on the sparsity α_n, since intuitively, if there is more data available where the communities are almost the same as the present time step, the density of edges should not need to be as large. We solve this in the following theorem, which is the central contribution of this paper.

Theorem 1. *Consider the deterministic DSBM with any B_0. Define*

$$\rho_n \overset{\mathrm{def.}}{=} \min\left(1, \sqrt{\bar{n}_{\max} \alpha_n \varepsilon_n}\right). \tag{16}$$

Assume $t \geqslant t_{\min} \overset{\mathrm{def.}}{=} \frac{\log\left(\frac{\rho_n}{\alpha_n n}\right)}{2\log(1-\rho_n)}$, and

$$\frac{\alpha_n}{\rho_n} \gtrsim \frac{\log n}{n}. \tag{17}$$

Consider either the uniform estimator $A_t^{\mathrm{smooth}} = A_t^{\mathrm{unif}}$ with $r \sim \frac{1}{\rho_n}$ or the exponential estimator $A_t^{\mathrm{smooth}} = A_t^{\mathrm{exp}}$ with $\lambda \sim \rho_n$.

For all $\nu > 0$, there is a universal constant C_ν such that, with probability at least $1 - n^{-\nu}$, it holds that

$$\|A_t^{\mathrm{smooth}} - P_t\| \leqslant C_\nu \sqrt{n\alpha_n \rho_n}. \tag{18}$$

This result is proved in section 6.2. In this theorem, we improve over [37] in several ways. First, we improve $\rho_n^{(\mathrm{PZ})}$ to ρ_n by replacing n with $\bar{n}_{\max} \leqslant n$. In the case where $\sum_\ell (B_0)_{k\ell}$ stays bounded, for instance if it is defined as (3) with $\tau \sim \frac{1}{K}$, we have $\bar{n}_{\max} \sim \frac{n}{K}$ and this improves the bound (18) compared to (14). We also extend the result to the exponential estimator with the right choice of forgetting factor. In fact, it can be seen in the proof that this result is valid for a more general class of estimators based on weighted averages to which the uniform and exponential estimators belong, see Section 6.

More importantly, the main feature of our result is the weaker condition (17), which relates the sparsity and the smoothness of the DSBM. Strinkingly, if

$$\varepsilon_n \sim \frac{n/\bar{n}_{\max}}{\log^2 n}, \tag{19}$$

which is a slight strengthening of (15), then our result is valid in the sparse regime $\alpha_n \sim \frac{1}{n}$, which is a significant improvement compared to previous works. In any case, if we have exactly $\frac{\alpha_n}{\rho_n} \sim \frac{\log n}{n}$, then as previously Lemma 1 yields that $E(\hat{\Theta}, \Theta) \to 0$ when $K = o(\sqrt{\log n})$. At the limit, when $\varepsilon_n \to 0$ (and the number of steps $t_{\min}$ grows accordingly), the results stay valid for even sparser graphs $\alpha_n \to 0$, as long as (17) holds.

Proposition 2. *The result of Theorem 1 stays valid under the Markov DSBM, by replacing $\bar{n}_{\max}$ with n everywhere, and ρ_n by*

$$\rho_n = \min(1, \max(\sqrt{n\alpha_n\varepsilon_n}, \sqrt{\alpha_n}(n\log n)^{1/4})). \tag{20}$$

In this case, "with probability at least $1 - n^{-\nu}$" refers to joint probability on both the A_t and the Θ_t.

The above result (proved in section 6.2.3) shows that the Markov DSBM yields the exact same error bounds than the deterministic DSBM model, but since we do not assume a maximal community size here, $\bar{n}_{\max}$ is replaced with n. Furthermore, it seems that the rate of the deterministic case is reached when $\varepsilon_n \gtrsim \sqrt{\log n/n}$, or, in other words, when n is sufficiently large when ε_n gets small. Indeed, as ε_n gets small, concentration bounds on the Markov chain need to hold over an increasingly large number of steps, and in results n needs to grow accordingly. Nevertheless, the condition $\varepsilon_n \gtrsim \sqrt{\log n/n}$ is much weaker than the rate (19) for instance, such that the sparse regime with sufficient smoothness is still valid. In this particular Markov model, it may still be possible to improve these results in the limit $\varepsilon_n \to 0$, which we leave for future investigations.

Remark 2. As already observed in [37], with this proof technique, a constant ε_n, or in other words, a fraction of changing nodes s that grows linearly with n, does not result in an improvement of the rate of the error bounds compared to the static case. Following the statistical physic approach in the sparse static case [21, 31, 1], a conjecture on the detectability threshold in the sparse case and $\varepsilon_n \sim 1$ has been formulated in [15], but the proof is still open. Note that, as mentioned before, even in the static case, this analysis does not cover classic SC algorithm, or the case $K > 2$.

5. Spectral concentration of the normalized Laplacian

As mentioned in the introduction, the spectral concentration of the normalized Laplacian has been less studied than the adjacency matrix, even in the static case. Many works study the asymptotic spectral convergence of the normalized Laplacian in the dense case [43], but few examine non-asymptotic bounds.

5.1. Static case

Among the few existing bounds, [34] proves a concentration in $\mathcal{O}(1)$ in the relatively sparse case, and [38] proves a concentation in Frobenius norm but with the stronger condition $\alpha_n \gtrsim \frac{1}{\sqrt{\log n}}$. An important corollary of our study of the dynamic case is to significantly improve over these results, and obtain, to our knowledge, the best bound available in the relatively sparse case. We state the following proposition for any Bernoulli matrix (not necessarily SBM).

Proposition 3 (Normalized Laplacian, static case)**.** *Let A be a symmetric matrix with independent entries $a_{ij} \sim Ber(p_{ij})$. Assume $p_{ij} \leqslant \alpha_n$, and that there*

is $\bar{n}_{\min}, \bar{n}_{\max}$ such that for all i, $\alpha_n \bar{n}_{\min} \leqslant \sum_j p_{ij} \leqslant \alpha_n \bar{n}_{\max}$, and $\mu_B = \frac{\bar{n}_{\max}}{\bar{n}_{\min}}$. For all $\nu > 0$, there are constants C_ν, C'_ν such that: if

$$\alpha_n \geqslant C'_\nu \mu_B \frac{\log n}{\bar{n}_{\min}}, \tag{21}$$

then with probability at least $1 - n^{-\nu}$ we have

$$\|L(A) - L(P)\| \lesssim \frac{C_\nu \mu_B \sqrt{n}}{\bar{n}_{\min}\sqrt{\alpha_n}}.$$

Proof. This is a direct consequence of Theorem 4 in Section 6. □

In other words, when $\bar{n}_{\min} \sim n$ (for instance when all the p_{ij}/α_n are bounded below), then in the relatively sparse case the spectral concentration of the normalized Laplacian is in $\frac{1}{\sqrt{\log n}}$, which is a strict improvement over existing bounds.

Let us comment a bit on the condition (21). When $\bar{n}_{\min} = o(n)$ or $\mu_B^{-1} = o(1)$, it is stronger than the relatively sparse case. The attentive reader would also remark the subtle interplay of the quantifiers with the rate ν: in the analysis of the adjacency matrix in the previous section, any multiplicative constant between α_n and $\frac{\log n}{n}$ was acceptable, and the rate ν only forced a multiplicative constant C_ν in the final error bound. Here, the rate ν also imposes a multiplicative constant C'_ν in the sparsity hypothesis, which is essential to avoid having too many nodes with small degrees [25, 7].

5.2. Dynamic case

To our knowledge, the normalized Laplacian in the DSBM has never been studied theoretically. Our result is the following.

Theorem 2. *Consider the deterministic DSBM with B satisfying* (3), *and either the uniform estimator* $A_t^{\mathrm{smooth}} = A_t^{\mathrm{unif}}$ *with* $r \sim \frac{1}{\rho_n}$ *or the exponential estimator* $A_t^{\mathrm{smooth}} = A_t^{\exp}$ *with* $\lambda = \rho_n$. *Assume* $t \geqslant t_{\min}$.
For all $\nu > 0$, *there exist universal constants* $C_\nu, C'_\nu > 0$ *such that: if*

$$\frac{\alpha_n}{\rho_n} \geqslant C'_\nu \mu_B \frac{\log n}{\bar{n}_{\min}}, \tag{22}$$

then with probability at least $1 - n^{-\nu}$, *it holds that*

$$\|L(A_t^{\mathrm{smooth}}) - L(P_t)\| \leqslant C_\nu \mu_B \sqrt{\frac{n\rho_n}{\bar{n}_{\min}^2 \alpha_n}}. \tag{23}$$

This result is proved in section 6.3. In the case of balanced communities, the result of Theorem 2 combined with Lemma 1 yields the same error rate than in the case of the adjacency matrix with Theorem 1 and Lemma 1, even in terms of K when $\bar{n}_{\min}, \bar{n}_{\max} \sim \frac{n}{K}$. Thus all the observations of the previous section

remain valid, including the fact that A^{smooth} may belong to a more general class of estimator (see Section 6). However in the Laplacian case, the condition (22) is slightly stronger than (17), similar to the static case. In practice though, it is well-known that the normalized Laplacian generally performs better (Fig. 1). This is not fully explained by our theory and left for future work. Finally, note that we only provide the result in the deterministic DSBM case, indeed, the use of the normalized Laplacian requires the existence of a minimal community size $\bar{n}_{\min}$, which is not always guaranteed in the Markov DSBM.

6. Proofs

In this section, we provide the proof of our main results, largely inspired by [25] and [37]. The technical computations are given in appendix. Despite some similarity with [25] and [37], we strove to make the proofs self-contained.

6.1. Preliminaries

We place ourselves at a particular time t. Both estimators A_t^{unif} and A_t^{exp} can be written as a weighted sum

$$A_t^{\text{smooth}} = \sum_{k=0}^{t} \beta_k A_{t-k} \,, \tag{24}$$

where $\beta_0 = \ldots = \beta_{r-1} = \frac{1}{r}$ and $\beta_k = 0$ for $k \geqslant r$ in the uniform case, and $\beta_k = \lambda(1-\lambda)^k$ for $k < t$ and $\beta_t = (1-\lambda)^t$ in the exponential case. As we will see, our results will be valid for any estimator of the form (24), with weights $\beta_k \geqslant 0$ that satisfy: there are constants $\beta_{\max}, C_\beta, C'_\beta > 0$ such that:

$$\sum_{k=0}^{t} \beta_k = 1, \quad \beta_k \leqslant \beta_{\max}, \quad \sum_{k=0}^{t} \beta_k^2 \leqslant C_\beta \beta_{\max}, \tag{25}$$
$$\sum_{k=0}^{t} \beta_k \min(1, \sqrt{k \varepsilon_n}) \leqslant C'_\beta \sqrt{\frac{\varepsilon_n}{\beta_{\max}}}.$$

In words, the weights must naturally sum to 1 and be bounded; the sum of their squares must be small; and they must decrease faster than $\sqrt{k}$, which is roughly the rate at which the past communities Θ_{t-k} deviate from Θ_t. It is not difficult to show that the uniform and exponential estimator satisfy these conditions.

Lemma 2. *The weights in the uniform estimator* (6) *satisfy* (25) *with* $\beta_{\max} = \frac{1}{r}$, $C_\beta = C'_\beta = 1$. *If* $t \geqslant t_{\min} = \frac{\min(\log(\varepsilon_n/\beta_{\max}), \log \beta_{\max})}{2 \log(1-\beta_{\max})}$, *the weights in the exponential estimator* (7) *satisfy* (25) *with* $\beta_{\max} = \lambda$, $C_\beta = \frac{3}{2}$, $C'_\beta = 2$.

Proof. The computations are trivial in the uniform case, where the last condition is implied by the stronger property $\sum_k \beta_k \sqrt{k} \leqslant \sqrt{r} = \sqrt{\frac{1}{\beta_{\max}}}$. In the exponential

case, we have $\beta_{\max} = \lambda$, and

$$\sum_k \beta_k^2 = \lambda^2 \sum_{k=0}^{t-1}(1-\lambda)^{2k} + (1-\lambda)^{2t} \leqslant \lambda^2 \sum_{k=0}^{\infty}(1-\lambda)^{2k} + \lambda \leqslant \frac{3}{2}\lambda,$$

where the first inequality is valid since $t \geqslant \frac{\log \beta_{\max}}{2\log(1-\beta_{\max})}$, and thus $C_\beta = \frac{3}{2}$. Next, we have

$$\sum_{k=0}^{t} \beta_k \min(1, \sqrt{k\varepsilon_n}) \leqslant \sqrt{\varepsilon_n}\lambda \sum_{k=0}^{\infty} \sqrt{k}(1-\lambda)^k + (1-\lambda)^t.$$

Since $t \geqslant \frac{\log(\varepsilon_n/\beta_{\max})}{2\log(1-\beta_{\max})}$, we get $(1-\lambda)^t \leqslant \sqrt{\frac{\varepsilon_n}{\lambda}}$ and

$$\sum_{k=0}^{\infty}(1-\lambda)^k\sqrt{k} \leqslant \sqrt{\sum_{k=0}^{\infty}(1-\lambda)^k}\sqrt{\sum_{k=0}^{\infty}(1-\lambda)^k k} = \sqrt{\frac{1}{\lambda}}\sqrt{\frac{1-\lambda}{\lambda^2}} \leqslant \frac{1}{\lambda^{3/2}}.$$

and therefore we obtain the desired inequality with $C_\beta' = 2$.   □

6.2. Concentration of adjacency matrix: proof of Theorem 1

For an estimator of the form (24), our goal is to bound $\|A_t^{\mathrm{smooth}} - P_t\|$. We define $P_t^{\mathrm{smooth}} \stackrel{\mathrm{def.}}{=} \sum_{k=0}^{t} \beta_k P_{t-k}$, and divide the error in two terms:

$$\|A_t^{\mathrm{smooth}} - P_t\| \leqslant \|A_t^{\mathrm{smooth}} - P_t^{\mathrm{smooth}}\| + \|P_t^{\mathrm{smooth}} - P_t\|. \qquad (26)$$

The first error term corresponds to the difference between A_t^{smooth} and its expectation (up to the diagonal terms). Intuitively, it *decreases* when the amount of smoothing *increases*, that is, when r increases or λ gets close to 0, since the sum of matrices is taken over more values. The second term is the difference between the smoothed matrix of probability connection and its value at time t. This time, it will increase when the amount of smoothing increases, since the past communities will be increasingly present in P_t^{smooth}. Once we have the two bounds, we can balance them to obtain an optimal value for r or λ, respectively $\frac{1}{\rho_n}$ and ρ_n.

6.2.1. Bound on the first term

The first bound will be handled by the following general concentration theorem. This is where we are able to weaken the hypothesis on the sparsity.

Theorem 3. *Let $A_1, \ldots, A_t \in \{0,1\}^{n \times n}$ be t symmetric Bernoulli matrices whose elements $a_{ij}^{(k)}$ are independent random variables:*

$$a_{ij}^{(k)} \sim Ber(p_{ij}^{(k)}), \quad a_{ji}^{(k)} = a_{ij}^{(k)}, \quad a_{ii}^{(k)} = 0.$$

Assume $p_{ij}^{(k)} \leqslant \alpha_n$. *Consider non-negative weights* β_k *that satisfy* (25). *Denoting* $A = \sum_{k=0}^t \beta_k A_{t-k}$ *and* $P = \mathbb{E}(A)$, *there is a universal constant* C *such that for all* $c > 0$ *we have*

$$\mathbb{P}\left(\|A - P\| \geqslant C(1+c)\sqrt{n\alpha_n\beta_{\max}}\right) \leqslant e^{-\left(\frac{c^2/2}{2C_\beta + \frac{2}{3}c} - \log(14)\right)n} \tag{27}$$

$$+ e^{-\frac{c^2/2}{C_\beta + 2c/3} \cdot \frac{n\alpha_n}{\beta_{\max}} + \log n} + n^{-\frac{c}{4} + 6}.$$

This theorem is proved in Appendix A.2. Its proof is heavily inspired by [25] and [37]: the spectral norm is expressed as a maximization problem over the sphere, and for each point of the sphere the obtained sum is divided into so-called "light" terms, for which Berstein's concentration inequality is sufficient, and more problematic "heavy" terms, that require a complex concentration method. We obtain our weaker sparsity hypothesis in a small but crucial part of this second step, the so-called *bounded degree lemma*.

Lemma 3 (Bounded degree.). *Denote* $d_{i,t} = \sum_j (A_t)_{ij}$ *the degree of node* i *at time* t, $d_i = \sum_{k=0}^t \beta_k d_{i,t-k}$ *the smoothed degree and* $\bar{d}_i = \mathbb{E}d_i$. *Then, for all* c,

$$\mathbb{P}(\max_i|d_i - \bar{d}_i| \geqslant cn\alpha_n) \leqslant \exp\left(-\frac{c^2/2}{C_\beta + 2c/3} \cdot \frac{n\alpha_n}{\beta_{\max}} + \log n\right).$$

Proof. We use Bernstein's inequality. For any fixed i we have

$$d_i = \sum_{k=0}^t \sum_{j \neq i} \beta_k a_{ij}^{(t-k)} = \sum_{k,j} Y_{jk}.$$

where $Y_{jk} = \beta_k a_{ij}^{(t-k)}$ are such that $\mathbb{E}(Y_{jk}) = \beta_k p_{ij}^{(t-k)} \leqslant \beta_k \alpha_n$, $|Y_{jk} - \mathbb{E}Y_{jk}| \leqslant (\alpha_n + 1)\beta_k \leqslant 2\beta_{\max}$, and $Var(Y_{jk}) \leqslant \beta_k^2 \alpha_n$ such that $\sum_{k,j} Var(Y_{jk}) \leqslant C_\beta n\alpha_n\beta_{\max}$.

Therefore, applying Berstein's inequality, we have

$$\mathbb{P}(|d_i - \bar{d}_i| \geqslant cn\alpha_n) \leqslant \exp\left(-\frac{c^2 n^2 \alpha_n^2/2}{C_\beta n\alpha_n\beta_{\max} + \frac{2}{3}\beta_{\max}cn\alpha_n}\right).$$

Applying a union bound over the nodes i proves the result. $\qquad\square$

In the static case [25] where $\beta_{\max} = 1$, the bounded degree lemma is exactly where the relative sparsity hypothesis $\alpha_n \gtrsim \frac{\log n}{n}$ is needed, otherwise the probability of failure diverges. In the dynamic case, we see that $\beta_{\max}$ (which we will ultimately set at ρ_n) intervenes and gives our final hypothesis on sparsity and smoothness.

Applying Theorem 3, we obtain that for any fixed $\Theta_0, \ldots, \Theta_t$, if $\frac{n\alpha_n}{\beta_{\max}} \gtrsim \log n$, then for any $\nu > 0$ there is a constant C_ν such that with probability at least $1 - n^{-\nu}$

$$\|A_t^{\text{smooth}} - P_t^{\text{smooth}}\| \leqslant \|A_t^{\text{smooth}} - \mathbb{E}(A_t^{\text{smooth}})\| + \|\text{diag}(P_t^{\text{smooth}})\|$$

$$\leqslant C_\nu \sqrt{n\alpha_n \beta_{\max}} + \alpha_n.$$

Since in all considered cases we will have $\beta_{\max} \geqslant 1/n$ the second term is negligible, and we obtain

$$\|A_t^{\mathrm{smooth}} - P_t^{\mathrm{smooth}}\| \lesssim \sqrt{n\alpha_n \beta_{\max}}. \tag{28}$$

6.2.2. Second term

The second error term in (26) is handled slightly differently in the deterministic and Markov DSBM, even if the final bound is the same.

Lemma 4. *Consider the deterministic DSBM, with weights that satisfy* (25). *It holds that*

$$\|P_t^{\mathrm{smooth}} - P_t\| \lesssim C_\beta' \alpha_n \sqrt{\frac{n\bar{n}_{\max}\varepsilon_n}{\beta_{\max}}}. \tag{29}$$

Proof. Since the weights sum to 1, we decompose

$$\|P_t^{\mathrm{smooth}} - P_t\| \leqslant \sum_k \beta_k \|P_{t-k} - P_t\| \leqslant \sum_k \beta_k \|P_{t-k} - P_t\|_F.$$

where $\|\cdot\|_F$ is the Frobenius norm. Consider $P = \Theta B \Theta^\top$ and $P' = \Theta' B (\Theta')^\top$ two probability matrices such that there is a set $\mathcal{S}$ of nodes that have changed communities. We have then:

$$\|P - P'\|_F^2 = \sum_{i \in \mathcal{S}} \sum_j (p_{ij} - p_{ij}')^2 + (p_{ji} - p_{ji}')^2 \leqslant 4 \sum_{i \in \mathcal{S}} \sum_j p_{ij}^2 + (p_{ij}')^2$$

$$\leqslant 8|\mathcal{S}|\alpha_n^2 \bar{n}_{\max}.$$

Since $\sum_j p_{ij}^2 \leqslant \alpha_n^2 \bar{n}_{\max}$ and at most ks nodes have changed community between P_t and P_{t-k}, with a maximum of n nodes, we have

$$\|P_{t-k} - P_t\|_F^2 \leqslant 2\alpha_n^2 \bar{n}_{\max} \min(n, ks) = 2\alpha_n^2 n \bar{n}_{\max} \min(1, k\varepsilon_n). \tag{30}$$

Using the hypothesis that we have made on $\sum_k \beta_k \min(1, \sqrt{k\varepsilon_n})$, we obtained the desired bound. $\qquad\square$

At the end of the day, combining (26), (28) and (29) for both deterministic and Markov DSBM model we obtain with the desired probability:

$$\|A_t^{\mathrm{smooth}} - P_t\| \lesssim E_1(\beta_{\max}) + E_2(\beta_{\max}) \text{ where } \begin{cases} E_1(\beta) \stackrel{\mathrm{def.}}{=} \sqrt{n\alpha_n \beta}, \\ E_2(\beta) \stackrel{\mathrm{def.}}{=} \alpha_n \sqrt{\frac{n\bar{n}_{\max}\varepsilon_n}{\beta}}. \end{cases} \tag{31}$$

As expected, E_1 decreases and E_2 increases when $\beta_{\max}$ decreases. A simple function study show that the sum of the errors is minimized for $\beta_{\max} = \rho_n$, which concludes the proof of Theorem 1.

6.2.3. Markov DSBM

Since the bound on the first term (28) is valid for any Θ_k, and the A_k are conditionally independent given the Θ_k, by the law of total probability it is also valid with joint probability at least $1 - n^{-\nu}$ on both the A_k and Θ_k in the Markov DSBM model. For the bound on the second term, we show that (30) is still valid with high probability, replacing $\bar{n}_{\max}$ with n.

Lemma 5. *Consider the Markov DSBM model. We have*

$$\mathbb{P}\left(\exists k, \ \|P_{t-k} - P_t\|_F^2 \geqslant 4\alpha_n^2 n^2 \min(1, k\bar{\varepsilon}_n)\right) \leqslant \frac{1}{\bar{\varepsilon}_n}e^{-2\bar{\varepsilon}_n^2 n}.$$

where $\bar{\varepsilon} \stackrel{\text{def.}}{=} \max(\varepsilon_n, \sqrt{\log n/n})$.

The proof is in Appendix A.4. Using this Lemma, if (20) is satisfied we obtain that with probability at least $1 - n^{-\nu}$, (30) is satisfied for all k. Using the rest of the proof of Lemma 4, (29) is valid in the Markov DSBM model, with n instead of $\bar{n}_{\max}$ and $\bar{\varepsilon}_n$ instead of ε_n. The rest of the proof is the same as the deterministic case.

6.3. Concentration of Laplacian: proof of Theorem 2

A crucial part of handling the normalized Laplacian is to lower-bound the degrees of the nodes, since we later manipulate the inverse of the degree matrix. Under our hypotheses, the minimal expected degree is of the order of $\alpha_n\bar{n}_{\min}$, so we need to bound the deviation of the degrees with respect to this quantity. We revisit the bounded degree lemma.

Lemma 6 (Bounded degree revisited.)**.** *Under the deterministic DSBM, for all* c,

$$\mathbb{P}(\max_i|d_i - \bar{d}_i| \geqslant c\bar{n}_{\min}\alpha_n) \leqslant \exp\left(-\frac{c^2/2}{C_\beta + 2c/3}\cdot\frac{\bar{n}_{\min}\alpha_n}{\mu_B\beta_{\max}} + \log n\right).$$

Proof. We do the same proof as Lemma 3, but we remark that $\sum_{k,j} Var(Y_{jk}) \leqslant C_\beta\bar{n}_{\max}\alpha_n\beta_{\max}$, since $\sum_i p_{ij}^{(t-k)} \leqslant \alpha_n\bar{n}_{\max}$ for all k, i. Therefore, by Berstein's inequality, we have

$$\mathbb{P}(|d_i - \bar{d}_i| \geqslant c\bar{n}_{\min}\alpha_n) \leqslant \exp\left(-\frac{c^2\bar{n}_{\min}^2\alpha_n^2/2}{C_\beta\bar{n}_{\max}\alpha_n\beta_{\max} + \frac{2}{3}\beta_{\max}c\bar{n}_{\max}\alpha_n}\right).$$

Applying a union bound over the nodes i proves the result. $\qquad\square$

To lower-bound d_i, we use Lemma 6 with $0 < c < 1$, for instance $c = \frac{1}{2}$. The sparsity hypothesis (22) in the theorem comes directly from this: it uses $\bar{n}_{\min}$ instead of n, and the multiplicative constant C'_ν actually depends on the desired concentration rate ν, unlike the previous case of the adjacency matrix where ν

could be obtained by adjusting c in Lemma 3. Let us now turn to the proof of the theorem.

As before, we divide the bound in two parts:

$$\|L(A_t^{\mathrm{smooth}}) - L(P_t)\| \leqslant \|L(A_t^{\mathrm{smooth}}) - L(P_t^{\mathrm{smooth}})\| + \|L(P_t^{\mathrm{smooth}}) - L(P_t)\|. \tag{32}$$

The first bound is handled with a general concentration theorem.

Theorem 4. *Let* $A_1, \ldots, A_t \in \{0,1\}^{n \times n}$ *be* t *symmetric Bernoulli matrices whose elements* $a_{ij}^{(k)}$ *are independent random variables:*

$$a_{ij}^{(k)} \sim Ber(p_{ij}^{(k)}), \quad a_{ji}^{(k)} = a_{ij}^{(k)}, \quad a_{ii}^{(k)} = 0.$$

Consider non-negative weights β_k *that satisfy* (25). *Denoting* $A = \sum_{k=0}^t \beta_k A_{t-k}$ *and* $P = \mathbb{E}(A)$. *Assume* $p_{ij}^{(k)} \leqslant \alpha_n$, *and that there is* $\bar{n}_{\min}, \bar{n}_{\max}$ *such that for all* i, $\alpha_n \bar{n}_{\min} \leqslant \sum_j p_{ij} \leqslant \alpha_n \bar{n}_{\max}$. *Then there is a universal constant* C *such that for all* $c > 0$ *we have*

$$\mathbb{P}\Big(\|L(A) - L(P)\| \geqslant \frac{C(1+c)\mu_B}{\bar{n}_{\min}} \sqrt{\frac{n\beta_{\max}}{\alpha_n}}\Big) \tag{33}$$

$$\leqslant e^{-\Big(\frac{c^2/2}{2C_\beta + \frac{2}{3}c} - \log(14)\Big)n} + e^{-\frac{c^2/2}{C_\beta + 2c/3} \cdot \frac{n\alpha_n}{\beta_{\max}} + \log n}$$

$$+ e^{-\frac{1/8}{C_\beta + 1/3} \cdot \frac{\bar{n}_{\min}\alpha_n}{\mu_B\beta_{\max}} + \log n} + n^{-\frac{c}{4} + 6}.$$

The proof is in Appendix A.3. Similar to the adjacency matrix case, we thus obtain

$$\|L(A_t^{\mathrm{smooth}}) - L(P_t^{\mathrm{smooth}})\| \leqslant \|L(A_t^{\mathrm{smooth}}) - L(\mathbb{E}(A_t^{\mathrm{smooth}}))\|$$
$$+ \|L(\mathbb{E}(A_t^{\mathrm{smooth}})) - L(P_t^{\mathrm{smooth}})\|,$$

and by Lemma 11, the second term is negligible since $\mathbb{E}(A_t^{\mathrm{smooth}})$ and P_t^{smooth} only differ by their diagonal, of the order of α_n.

The second bound is handled in the same way as the adjacency matrix in the deterministic case.

Lemma 7. *Under the deterministic DSBM, we have*

$$\|L(P_t) - L(P_t^{\mathrm{smooth}})\| \lesssim \frac{C_\beta' \mu_B}{\bar{n}_{\min}} \sqrt{\frac{n\bar{n}_{\max}\varepsilon_n}{\beta_{\max}}}.$$

The proof is in Appendix A.4.

At the end of the day, we obtain

$$\|L(A_t^{\mathrm{smooth}}) - L(P_t)\| \lesssim \frac{\mu_B}{\bar{n}_{\min}\alpha_n}(E_1(\beta_{\max}) + E_2(\beta_{\max})), \tag{34}$$

which is minimized for the same choice of $\beta_{\max} \sim \rho_n$.

7. Conclusion and outlooks

In the DSBM, it should come as no surprise that a model that is very smooth should not need to be as dense as when treading with a single snapshot. Our analysis is the first to show this, for classic SC, in a non-asymptotic manner. Under a slightly stronger condition on the smoothness than that in [37], we showed that strong consistency guarantees can be obtained even in the sparse case. We extended the results to the normalized Laplacian and, although we obtain the same final error rate as the adjacency matrix, our analysis also yields, to our knowledge, the best non-asymptotic spectral bound concentration of the normalized Laplacian for Bernoulli matrices with independent edges.

In this theoretical paper, we did not discuss how to select in practice the various parameters of the algorithms such as the number of communities K, the forgetting factor λ, or the window size r. This is left for future investigations, as well as the analysis of varying K, n, or B. As we mentioned in Remark 2, an outstanding conjecture about the sparse case and $\varepsilon_n \sim 1$ is formulated in [15]. Finally, our new spectral concentration of the normalized Laplacian, which shows that $\|L(A) - L(P)\| \to 0$ in the relatively sparse case, may have consequences in other asymptotic analyses of the spectral convergence of the normalized Laplacian [43, 40, 26].

Appendix A: Proofs

A.1. Proof of Lemma 1

From [25, Section 5.4], for *any* matrix $M \in \mathbb{R}^{K \times K}$ and $Q = \Theta M \Theta^\top$, given an estimator $\hat{Q}$ that we feed to the SC algorithm it holds that

$$E(\hat{\Theta}, \Theta) \lesssim (1 + \delta) \frac{n'_{\max} K}{n n^2_{\min} \gamma^2_M} \|\hat{Q} - Q\|^2,$$

where γ_M is the smallest eigenvalue of M.

When using the adjacency matrix $\hat{P} = A$ to estimate the probability matrix $P = \Theta B \Theta^\top$, we have $B = \alpha_n B_0$, and $\gamma_M = \alpha_n \gamma$, which gives us (8). When B_0 is defined as (3), we have $\gamma = 1 - \tau$.

In the Laplacian case, for a node $i \leqslant n$ belonging to a community $k \leqslant K$, we have $d_i = \sum_j p_{ij} = d'_k \stackrel{\text{def.}}{=} (n_k - 1) B_{kk} + \sum_{\ell \neq k} n_\ell B_{k\ell} \leqslant \alpha_n \bar{n}_{\max}$, hence the Laplacian of the probability matrix $L(P)$ can be written as:

$$L(P) = D(P)^{-\frac{1}{2}} P D(P)^{-\frac{1}{2}} = D(P)^{-\frac{1}{2}} \Theta B \Theta^\top D(P)^{-\frac{1}{2}} = \Theta \left(D_B^{-\frac{1}{2}} B D_B^{-\frac{1}{2}} \right) \Theta^\top,$$

where $D_B = \text{diag}(d'_k) \in \mathbb{R}^{K \times K}$. Hence we can apply the result above with $M = D_B^{-\frac{1}{2}} B D_B^{-\frac{1}{2}}$, and since

$$\|D_B^{\frac{1}{2}} B^{-1} D_B^{\frac{1}{2}}\| \leqslant \frac{\alpha_n \bar{n}_{\max}}{\alpha_n \gamma} \leqslant \frac{\bar{n}_{\max}}{\gamma},$$

the smallest eigenvalue of $D_B^{-\frac{1}{2}} B D_B^{-\frac{1}{2}}$ satisfies $\gamma_M \geqslant \frac{\gamma}{\bar{n}_{\max}}$, which leads to the result.

A.2. Proof of Theorem 3

The proof is heavily inspired by [25]. Define $P_k = \mathbb{E}(A_k)$, $W_k = A_k - P_k$ and $w_{ij}^{(k)}$ its elements, and their respective smoothed versions $A = \sum_{k=0}^{t} \beta_k A_{t-k}$, $P = \mathbb{E}(A)$, $W = A - P$, and a_{ij}, p_{ij}, w_{ij} their elements. Denote by S the Euclidean ball in $\mathbb{R}^n$ of radius 1. The proof strategy of [25] is to define a grid

$$T = \{x \in S : 2\sqrt{n}x_i \in \mathbb{Z}\},$$

and simply note that (Lemma 2.1 in [25] supplementary):

$$\|W\| = \sup_{u \in S} |u^\top W u| \leqslant 4 \sup_{x,y \in T} |x^\top W y|.$$

Hence we must bound this last quantity. To do this, for each given (x, y) in T, we divide their indices into "light" pairs:

$$\mathcal{L}(x, y) = \{(i, j) : |x_i y_j| \leqslant \sqrt{\frac{\alpha_n}{\beta_{\max} n}}\},$$

and "heavy" pairs $\mathcal{H}(x, y)$ are all the other indices. We naturally divide

$$\sup_{x,y \in T} |x^\top W y| \leqslant \sup_{x,y \in T} |\sum_{(i,j) \in \mathcal{L}(x,y)} x_i y_j w_{ij}| + \sup_{x,y \in T} |\sum_{(i,j) \in \mathcal{H}(x,y)} x_i y_j w_{ij}|,$$

and bound each of these two terms separately.

A.2.1. Bounding the light pairs

To bound the light pairs, Bernstein's concentration inequality is sufficient.

Lemma 8 (Bounding the light pairs). *We have*

$$\mathbb{P}\left(\sup_{x,y \in T} |\sum_{(i,j) \in \mathcal{L}(x,y)} x_i y_j w_{ij}| \geqslant c\sqrt{n\alpha_n \beta_{\max}}\right) \leqslant 2e^{-\left(\frac{c^2/2}{2C_\beta + \frac{2}{3}c} - \log(14)\right)n}$$

for all constants $c > 0$.

Proof. The proof is immediate by applying Bernstein's inequality. Take any $(x, y) \in T$, denote $C = \sqrt{\frac{\alpha_n}{\beta_{\max} n}}$. Define $u_{ij} = x_i y_j 1_{(i,j) \in \mathcal{L}(x,y)} + x_j y_i 1_{(j,i) \in \mathcal{L}(x,y)}$ (which is necessary because the edges (i, j) and (j, i) are not independent). We have

$$\sum_{(i,j) \in \mathcal{L}(x,y)} x_i y_j w_{ij} = \sum_{1 \leqslant i < j \leqslant n} u_{ij} w_{ij} = \sum_{1 \leqslant i < j \leqslant n} \sum_{k=0}^{t} u_{ij} \beta_k w_{ij}^{(t-k)},$$

where $w_{ij}^{(t-k)}$ is a centered Bernoulli variable of parameter $p_{ij}^{(t-k)}$. Hence for all $i < j$, $0 \leqslant k \leqslant t$, let $Y_{ijk} = \beta_k u_{ij} w_{ij}^{(t-k)}$ be independent random variables such that $\mathbb{E} Y_{ijk} = 0$, $|Y_{ijk}| \leqslant 2C\beta_{\max}$ since $u_{ij} \leqslant 2C$, $\beta_k \leqslant \beta_{\max}$ and $w_{ij}^{(t-k)} \leqslant 1$, and $\sigma_{ijk}^2 := \mathrm{Var}(Y_{ijk}) \leqslant 2\beta_k^2(x_i^2 y_j^2 + x_j^2 y_i^2)\alpha_n$ since $\mathrm{Var}(w_{ij}^{(t-k)}) = p_{ij}^{(t-k)}(1 - p_{ij}^{(t-k)}) \leqslant \alpha_n$. Note that

$$\sum_{i<j}\sum_k \sigma_{ijk}^2 \leqslant 2\alpha_n \left(\sum_{i<j} x_i^2 y_j^2 + x_j^2 y_i^2\right) \cdot \left(\sum_{k=0}^t \beta_k^2\right)$$

$$\leqslant 2\alpha_n \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{j=1}^n y_j^2\right) C_\beta \beta_{\max}$$

$$\leqslant 2C_\beta \alpha_n \beta_{\max}.$$

Hence, applying Bernstein's inequality:

$$\mathbb{P}\left(|\sum_{(i,j)\in\mathcal{L}(x,y)} x_i y_j w_{ij}| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2/2}{2C_\beta \alpha_n \beta_{\max} + \frac{2}{3}C\beta_{\max}t}\right)$$

$$\mathbb{P}\left(|\sum_{(i,j)\in\mathcal{L}(x,y)} x_i y_j w_{ij}| \geqslant c\sqrt{\alpha_n\beta_{\max}n}\right) \leqslant 2\exp\left(-\left(\frac{c^2/2}{2C_\beta + \frac{2}{3}c}\right)n\right).$$

Then, we use the fact that $|T| \leqslant e^{n\log(14)}$ (see proof of Lemma 3.1 in [25]) and a union bound to conclude.   □

A.2.2. Bounding the heavy pairs

To bound the heavy pairs, two main Lemmas are required: the so-called *bounded degree* (Lemma 3) and *bounded discrepancy* lemma, presented below. As mentioned before, the bounded degree lemma is key in improving the sparsity hypothesis, despite the simplicity of its proof. The bounded discrepancy lemma is closer to its original proof [25], that we reproduce here for completeness.

Lemma 9 (Bounded discrepancy). *For $I, J \subset \{1, \ldots, n\}$, we define*

$$\mu(I, J) = \alpha_n |I||J|, \quad e(I, J) = \sum_{k=0}^t \beta_k e_{t-k}(I, J),$$

where $e_t(I, J)$ is the number of edges between I and J at time t. Then, for all c, c', with probability $1 - e^{-\frac{c^2/2}{C_\beta + 2c/3} \cdot \frac{n\alpha_n}{\beta_{\max}} + \log n} - n^{-\frac{c'}{4} + 6}$: for all $|I| \leqslant |J|$ at least one the following is true:

    1. $e(I, J) \leqslant c''\mu(I, J)$ with $c'' = \max(ec, 8)$;
    2. $e(I, J)\log\frac{e(I,J)}{\mu(I,J)} \leqslant c'\beta_{\max}|J|\log\frac{n}{|J|}$.

Of course by symmetry it is also valid for $|J| \leqslant |I|$ with the same probability (inverting the role of I and J in the bounds).

To prove it, we need the following Lemma.

Lemma 10 (Adapted from Lemma 9 in [37])**.** *Let $X_1, ..., X_n$ be independent variables such that $X_i = Y_i - \mathbb{E}Y_i$, where Y_i is a Bernoulli random variable with parameter p_i. Define $X = \sum_i w_i X_i$ where $0 \leqslant w_i \leqslant w_{\max}$. Let μ be such that $\sum_{i=1} w_i p_i \leqslant \mu$. Then, for all $t \geqslant 7$, we have*

$$\mathbb{P}\left(X \geqslant t\mu\right) \leqslant \exp\left(-\frac{t\log(1+t)\mu}{2w_{\max}}\right).$$

Proof. For some $\lambda > 0$ to be fixed later, have $\mathbb{E}(e^{\lambda w_i X_i}) = p_i e^{w_i(1-p_i)\lambda} + (1 - p_i)e^{w_i p_i \lambda}$. Hence

$$\mathbb{E}(e^{\lambda X}) = \prod_i \mathbb{E}(e^{\lambda w_i X_i}) = \prod_i \left(p_i e^{w_i(1-p_i)\lambda} + (1 - p_i)e^{w_i p_i \lambda}\right)$$

$$\leqslant e^{-\lambda \sum_i w_i p_i} \prod_i \left(1 + p_i(e^{w_i \lambda} - 1)\right).$$

Using $1 + a \leqslant e^a$ and $e^x - 1 \leqslant \frac{e^A - 1}{A}x$ for $0 \leqslant x \leqslant A$, we have

$$\prod_i \left(1 + p_i(e^{w_i \lambda} - 1)\right) \leqslant \exp\left(\frac{e^{w_{\max}\lambda} - 1}{w_{\max}} \sum_i w_i p_i\right).$$

Hence, for $t \geqslant 7$ and $\lambda = \frac{\log(1+t)}{w_{\max}}$,

$$\mathbb{P}(X \geqslant t\mu) \leqslant e^{-t\mu\lambda}\mathbb{E}(e^{\lambda X}) \leqslant \exp\left(\left(\frac{e^{w_{\max}\lambda} - 1}{w_{\max}} - \lambda\right)\sum_i w_i p_i - \lambda t\mu\right)$$

$$= \exp\left(\frac{1}{w_{\max}}\left((t - \log(1+t))\sum_i w_i p_i - \log(1+t)t\mu\right)\right).$$

Since $\log(1 + t) \leqslant t$ and $\sum_i w_i p_i \leqslant \mu$,

$$\mathbb{P}(X \geqslant t\mu) \leqslant \exp\left(\frac{\mu}{w_{\max}}(t - \log(1+t) - \log(1+t)t)\right)$$

$$\leqslant \exp\left(\frac{\mu}{2w_{\max}}t\log(1+t)\right)$$

since $t - \log(1 + t) \leqslant \frac{1}{2}t\log(1 + t)$. $\square$

The Lemma above is slightly stronger than Bernstein in this particular case: we would have obtained $O(t)$ instead of $O(t\log(1 + t))$. Now we can prove the bounded discrepancy lemma.

Proof of Lemma 9. We assume that the bounded degree property (Lemma 3) holds, which implies that for all I, J, it hold that:

$$e(I, J) = \sum_{k=0}^{t} \beta_k e_{t-k}(I, J) \leqslant \sum_k \beta_k \min \left(\sum_{i \in I} d_{i, t-k}, \ \sum_{j \in J} d_{j, t-k} \right)$$

$$\leqslant \min \left(\sum_{i \in I} d_i, \ \sum_{j \in J} d_j \right) \leqslant c n \alpha_n \min(|I|, |J|).$$

Following this, for any pair I, J such that $|I| \geqslant n/e$ or $|J| \geqslant n/e$ (where $e = \exp(1)$ is chosen for later conveniency), then $\frac{e(I,J)}{\mu(I,J)} \leqslant \frac{c n \alpha_n \min(|I|, |J|)}{\alpha_n |I| |J|} \leqslant ce$ and the result is proved.

Thus we now considers the pairs I, J where both have size less than n/e, and such that $|I| \leqslant |J|$ without lost of generality. For such a given pair I, J, we decompose

$$e(I, J) = \sum_{k=0}^{t} \sum_{(i,j)} \beta_k a_{ij}^{(t-k)} = \sum_{i,j,k} Y_{ijk},$$

where the sum over (i, j) counts only once each distinct edge between I and J, and $Y_{ijk} = a_{ij}^{(t-k)}$ is a Bernoulli variable with parameter $p_{ij}^{(t-k)}$. Using $\sum_{i,j,k} \beta_k p_{ij}^{(t-k)} \leqslant |I||J|\alpha_n = \mu(I, J)$ and Lemma 10, we have, for any $t \geqslant 8$,

$$\mathbb{P}(e(I, J) \geqslant t\mu(I, J)) \leqslant \mathbb{P}\left(e(I, J) - \mathbb{E}e(I, J) \geqslant (t - 1)\mu(I, J)\right)$$

$$\leqslant \exp\left(-\frac{\mu(I, J)(t - 1) \log k}{2\beta_{\max}}\right) \leqslant \exp\left(-\frac{\mu(I, J) t \log t}{4\beta_{\max}}\right).$$

Denoting $u = u(I, J)$ the unique value such that $u \log u = \frac{c' \beta_{\max} |J|}{\mu(I, J)} \log \frac{n}{|J|}$ and $t(I, J) = \max(8, u(I, J))$, we have (again for a fixed pair I, J of size less that n/e):

$$\mathbb{P}(e(I, J) \geqslant t(I, J)\mu(I, J)) \leqslant e^{-\frac{c'}{4}|J| \log \frac{n}{|J|}}.$$

Then, performing the same computations as in [25] (reproduced here for completeness):

$$\mathbb{P}\left(\exists I, J : |I| \leqslant |J| \leqslant \frac{n}{e}, \ e(I, J) \geqslant t(I, J)\mu(I, J)\right)$$

$$\leqslant \sum_{1 \leqslant |I| \leqslant |J| \leqslant n/e} e^{-\frac{c'}{4}|J| \log \frac{n}{|J|}} = \sum_{1 \leqslant h \leqslant g \leqslant n/e} \binom{n}{h}\binom{n}{g} e^{-\frac{c'}{4} g \log \frac{n}{g}}$$

$$\leqslant \sum_{1 \leqslant h \leqslant g \leqslant n/e} \left(\frac{ne}{h}\right)^h \left(\frac{ne}{g}\right)^g e^{-\frac{c'}{4} g \log \frac{n}{g}}$$

$$\leqslant \sum_{1 \leqslant h \leqslant g \leqslant n/e} \exp\left(h \log \frac{ne}{h} + g \log \frac{ne}{g} - \frac{c'}{4} g \log \frac{n}{g}\right)$$

$$\leqslant \sum_{1\leqslant h\leqslant g\leqslant n/e} \exp\left(4g\log\frac{n}{g} - \frac{c'}{4}g\log\frac{n}{g}\right) \leqslant \sum_{1\leqslant h\leqslant g\leqslant n/e} n^{-\frac{c'}{4}+4} \leqslant n^{-\frac{c'}{4}+6}$$

using the fact that $x \to x\log x$ is increasing on $[1, n/e]$. So, $e(I, J) \geqslant t(I, J)$ $\mu(I, J)$ holds uniformly for all pairs I, J with high probability.

Finally, we distinguish two cases depending on the value of $t(I, J)$. If $t(I, J) = 8$ we get $e(I, J) \leqslant 8\mu(I, J)$. If $t(I, J) = u(I, J) \geqslant 8$, we have $e(I, J) \leqslant u\mu(I, J)$, and

$$\frac{e(I, J)}{\mu(I, J)}\log\frac{e(I, J)}{\mu(I, J)} \leqslant u\log u \leqslant \frac{c'\beta_{\max}}{\mu(I, J)}|J|\log\frac{n}{|J|}. \qquad \square$$

We can now prove the bound on the heavy pairs, that is, we want to prove with high probability:

$$\sup_{x,y\in T}\left|\sum_{(i,j)\in\mathcal{H}(x,y)} x_i y_j w_{ij}\right| \lesssim \sqrt{n\alpha_n\beta_{\max}}.$$

Since $p_{ij} \leqslant \alpha_n$ and by definition of the heavy pairs, for all $x, y \in T$:

$$\left|\sum_{(i,j)\in\mathcal{H}(x,y)} x_i y_j p_{ij}\right| \leqslant \alpha_n \sum_{(i,j)\in\mathcal{H}(x,y)} \frac{x_i^2 y_j^2}{|x_i y_j|}$$

$$\leqslant \alpha_n\sqrt{\frac{n\beta_{\max}}{\alpha_n}}\|x\|^2\|y\|^2 \leqslant \sqrt{n\alpha_n\beta_{\max}}.$$

Hence our goal is now to bound $\sup_{x,y\in T}\left|\sum_{(i,j)\in\mathcal{H}(x,y)} x_i y_j a_{ij}\right|$. We will show that, when the bounded degree and bounded discrepancy properties hold, this sum is bounded for all x, y. From now on, we assume that these results hold, and consider any $x, y \in T$. Let us define sets of indices I_s, J_t over which we bound uniformly x_i and y_j, and replace the sum over a_{ij} is these sets by $e(I_s, J_t)$. More specifically, we define

$$I_s = \left\{i : 2^{s-1}\frac{1/2}{\sqrt{n}} \leqslant |x_i| < 2^s\frac{1/2}{\sqrt{n}}\right\} \quad \text{for} \quad s = 1, \dots, \log_2(2\sqrt{n}) + 1,$$

$$J_t = \left\{j : 2^{t-1}\frac{1/2}{\sqrt{n}} \leqslant |y_j| < 2^t\frac{1/2}{\sqrt{n}}\right\} \quad \text{for} \quad t = 1, \dots, \log_2(2\sqrt{n}) + 1.$$

Since we consider heavy pairs, we need only consider indices (s, t) such that $2^{s+t} \geqslant 16\sqrt{\frac{\alpha_n n}{\beta_{\max}}}$, and we define $C_n = \sqrt{\frac{\alpha_n n}{\beta_{\max}}}$ for convenience. Moreover, we have $\sum_{i\in I_s, j\in J_t} a_{ij} \leqslant 2e(I_s, J_t)$, since each edge indices appears at most twice. Hence, we have:

$$\left|\sum_{(i,j)\in\mathcal{H}(x,y)} x_i y_j a_{ij}\right| \leqslant \sum_{(s,t):2^{s+t}\geqslant 16C_n} 2^{s+t}\frac{1/4}{n} \cdot 2e(I_s, J_t). \qquad (35)$$

We now introduce more notations. We denote $\mu_s = \frac{4^s|I_s|}{n}, \nu_t = \frac{4^t|J_t|}{n}, \gamma_{st} = \frac{e(I_s,J_t)}{\alpha_n|I_s||J_t|}, \sigma_{st} = \gamma_{st}C_n2^{-(s+t)}$. We reformulate (35) as:

$$\left| \sum_{(i,j)\in\mathcal{H}(x,y)} x_iy_ja_{ij} \right| \leqslant \sum_{(s,t):2^{s+t}\geqslant 16C_n} 2^{s+t}\frac{1/2}{n} \cdot e(I_s,J_t)$$

$$= \frac{1}{2}\sqrt{n\alpha_n\beta_{\max}}\sum_{s,t} \frac{1}{\sqrt{n\alpha_n\beta_{\max}}}2^{-(s+t)}4^s4^t\frac{1}{n} \cdot \frac{e(I_s,J_t)\alpha_n|I_s||J_t|}{\alpha_n|I_s||J_t|} \cdot \frac{\sqrt{n}}{\sqrt{n}}$$

$$= \frac{1}{2}\sqrt{n\alpha_n\beta_{\max}}\sum_{s,t} \mu_s\nu_t\sigma_{st}. \tag{36}$$

Our goal is therefore to show that $\sum_{s,t}\mu_s\nu_t\sigma_{st} \lesssim 1$. For this, we will make extensive use of the fact that $\mu_s \leqslant 16\sum_{i\in I_s}x_i^2$, and therefore $\sum_s\mu_s \leqslant 16$, and similarly $\sum_t\nu_t \leqslant 16$.

Following the original proof of [25], let $\mathcal{C} = \{(s,t) : 2^{s+t} \geqslant 16C_n, |I_s| \leqslant |J_t|\}$, divided in six:

$$\mathcal{C}_1 = \{(s,t)\in\mathcal{C} : \sigma_{st} \leqslant 1\},$$
$$\mathcal{C}_2 = \{(s,t)\in\mathcal{C}\setminus\mathcal{C}_1 : \gamma_{st} \leqslant c_2\},$$
$$\mathcal{C}_3 = \{(s,t)\in\mathcal{C}\setminus(\cup_{i=1}^2\mathcal{C}_i) : 2^{s-t} \geqslant C_n\},$$
$$\mathcal{C}_4 = \{(s,t)\in\mathcal{C}\setminus(\cup_{i=1}^3\mathcal{C}_i) : \log\gamma_{st} > \frac{1}{4}\log\frac{4^t}{\nu_t}\},$$
$$\mathcal{C}_5 = \{(s,t)\in\mathcal{C}\setminus(\cup_{i=1}^4\mathcal{C}_i) : 2t\log 2 \geqslant \log\frac{1}{\nu_t}\},$$
$$\mathcal{C}_6 = \{(s,t)\in\mathcal{C}\setminus(\cup_{i=1}^5\mathcal{C}_i)\}.$$

Similarly, we define $\mathcal{C}' = \{(s,t) : 2^{s+t} \geqslant 16C_n, |I_s| \geqslant |J_t|\}$ and $\mathcal{C}_i'$ the same way by inverting the roles of μ_s and ν_t. We write the proof for $\mathcal{C}$, the other case is strictly symmetric. Our goal is to prove that each of the $\sum_{(s,t)\in\mathcal{C}_i}\mu_s\nu_t\sigma_{st}$ is bounded by a constant.

Pairs in $\mathcal{C}_1$ In this case we get

$$\sum_{(s,t)\in\mathcal{C}_1}\mu_s\nu_t\sigma_{st} \leqslant \sum_{s,t}\mu_s\nu_t \leqslant 16^2.$$

Pairs in $\mathcal{C}_2$ This includes the indices for which the first case in the bounded discrepancy lemma (Lemma 9) is satisfied. Since for $s,t\in\mathcal{C}$ we have $2^{s+t} \geqslant 16C_n$, we have $\sigma_{st} \leqslant \gamma_{st}/16$, and

$$\sum_{(s,t)\in\mathcal{C}_2}\mu_s\nu_t\sigma_{st} \leqslant c''\sum_{s,t}\mu_s\nu_t/16 \leqslant c''16.$$

Pairs in $\mathcal{C}_3$. Since $2^{s-t} \geqslant C_n$, we have necessarily $t \leqslant s - \log_2 C_n$. Furthermore, since we assumed the bounded degree property (Lemma 3), we have

$e(I_s, J_t) \leqslant c|I_s|\alpha_n n$, and therefore $\gamma_{st} \leqslant cn/|J_t|$. Then,

$$\sum_{(s,t)\in\mathcal{C}_3} \mu_s\nu_t\sigma_{st} \leqslant \sum_s \mu_s \sum_{t=1}^{s-\log_2 C_n} \frac{4^t|J_t|}{n}\frac{cn}{|J_t|}C_n 2^{-(s+t)}$$

$$= c\sum_s \mu_s C_n 2^{-s}\sum_{t=1}^{s-\log_2 C_n} 2^t$$

$$= c\sum_s \mu_s C_n 2^{-s}(2^{s-\log_2 C_n+1}-1)$$

$$\leqslant 2c\sum_s \mu_s \leqslant 32c.$$

Pairs in $\mathcal{C}_4$ All I_s, J_t for which the first case of the bounded discrepancy lemma is satisfied are included in $\mathcal{C}_2$, hence the remaining sets satisfy the second case of the lemma, which reads $e(I_s, J_t)\log\gamma_{st} \leqslant c'\beta_{\max}|J_t|\log\frac{n}{|J_t|}$. It can be reformulated as

$$\gamma_{st}\alpha_n|I_s||J_t|\log\gamma_{st} \leqslant c'\beta_{\max}|J_t|\log\frac{4^t}{\nu_t},$$

$$\sigma_{st}2^{s+t}\sqrt{\frac{\beta_{\max}}{\alpha_n n}}\alpha_n\mu_s 4^{-s}n\log\gamma_{st} \leqslant c'\beta_{\max}\log\frac{4^t}{\nu_t},$$

$$\sigma_{st}\mu_s\log\gamma_{st} \leqslant c'\frac{2^{s-t}}{C_n}\log\frac{4^t}{\nu_t}. \tag{37}$$

Since $(s,t)\notin\mathcal{C}_3$, we have $2^{s-t} \leqslant C_n$, and therefore $s \leqslant t+\log_2 C_n$. Since $(s,t)\in\mathcal{C}_4$ we have $\log\gamma_{st} \geqslant \frac{1}{4}\log\frac{4^t}{\nu_t}$, and (37) implies $\sigma_{st}\mu_s \leqslant 4c'\frac{2^{s-t}}{C_n}$. Then,

$$\sum_{(s,t)\in\mathcal{C}_4} \mu_s\nu_t\sigma_{st} \leqslant \sum_t \nu_t \sum_{s=1}^{t+\log_2 C_n} 4c'\frac{2^{s-t}}{C_n}$$

$$\leqslant 4c'\sum_t \nu_s\frac{2^{-t}}{C_n}(2^{t+\log_2 C_n+1}-1) \leqslant 128c'.$$

Pairs in $\mathcal{C}_5$ We have $\frac{1}{\nu_t} \leqslant 4^t$, and since $(s,t)\notin\mathcal{C}_4$, we have $\log\gamma_{st} \leqslant \frac{1}{4}\log\frac{4^t}{\nu_t} \leqslant t\log 2$ and $\gamma_{st} \leqslant 2^t$. On the other hand, since $(s,t)\notin\mathcal{C}_1$, $1 \leqslant \sigma_{st} = \gamma_{st}C_n 2^{-(s+t)} \leqslant C_n 2^{-s}$, and $s \leqslant \log_2 C_n$.

Because $(s,t)\notin\mathcal{C}_2$ and $c'' \geqslant 8$, $\log\gamma_{st} \geqslant 3\log 2$, combining with $\frac{1}{\nu_t} \leqslant 4^t$, equation (37) becomes:

$$\sigma_{st}\mu_s \leqslant c'\frac{2^{s-t}}{C_n}\frac{4t}{3} \leqslant c'\frac{2^{s+1}}{3C_n},$$

since $t2^{-t} \leqslant 1/2$.

Combining these two facts,

$$\sum_{(s,t)\in\mathcal{C}_5} \mu_s\nu_t\sigma_{st} \leqslant \sum_t \nu_t \sum_{s=1}^{\log_2 C_n} \frac{2}{3C_n}c'2^s \leqslant c'\sum_t \nu_s \frac{2}{3C_n}(2^{\log_2 C_n+1}-1) \leqslant \frac{64}{3}c'.$$

Pairs in $\mathcal{C}_6$ Finally, we have $0 \leqslant \log\gamma_{st} \leqslant \frac{1}{2}\log\frac{1}{\nu_t} \leqslant \log\frac{1}{\nu_t}$ because, respectively, $(s,t) \notin \mathcal{C}_2$, $(s,t) \notin \mathcal{C}_4$ and $(s,t) \notin \mathcal{C}_5$, so $\gamma_{st} \leqslant \frac{1}{\nu_t}$. Since by definition $2^{s+t} \geqslant C_n$, we have $t \geqslant \log_2 C_n - s$, and

$$\begin{aligned}
\sum_{(s,t)\in\mathcal{C}_6} \mu_s\nu_t\sigma_{st} &= \sum_{(s,t)\in\mathcal{C}_6} \mu_s\nu_t\gamma_{st}2^{-(s+t)}C_n \\
&\leqslant \sum_s \mu_s C_n 2^{-s} \sum_{t\geqslant\log_2 C_n-s} (1/2)^t \\
&= \sum_s \mu_s C_n 2^{-s}\left(2 - \frac{1-(1/2)^{\log_2 C_n-s}}{1-(1/2)}\right) = 2\sum_s \mu_s \leqslant 32.
\end{aligned}$$

We conclude the proof of Theorem 3 by gathering the bounds on the light and heavy pairs, with the corresponding probabilities of failure. We consider the same constant $c > 0$ in each lemma for simplicity.

A.3. Concentration of Laplacian: proof of Theorem 4

We note the degree matrices of A and P respectively D and D_P, containing the degrees $d_i = \sum_{kj} \beta_k a_{ij}^{(t-k)}$ and $\bar{d}_i = \mathbb{E}d_i$. Note that under our assumptions $d_{\min} \overset{\text{def.}}{=} \alpha_n\bar{n}_{\min} \leqslant \bar{d}_i \leqslant d_{\max} \overset{\text{def.}}{=} \alpha_n\bar{n}_{\max}$. Applying Lemma 6 with $c = \frac{1}{2}$, we obtain: for all $\nu > 0$, there is a constant C'_ν such that, if $\frac{\alpha_n}{\beta_{\max}} \geqslant C'_\nu\mu_B\frac{\log n}{\bar{n}_{\min}}$, with probability at least $1 - n^{-\nu}$ we have $\frac{1}{2}d_{\min} \leqslant d_i \leqslant \frac{3}{2}d_{\max}$ for all i. We assume that it is satisfied for the rest of the proof.

We apply Lemma 11, from which

$$\|L(A) - L(P)\| \leqslant \frac{2\|A-P\|}{d_{\min}} + \frac{4\|(D-D_P)P\|}{d_{\min}^2}. \tag{38}$$

We will now bound $\|A - P\|$ and $\|(D-D_P)P\|$ with high probability, and use a union bound to conclude.

By Theorem 3, with probability $1 - n^{-\nu}$ we have $\|A - P\| \lesssim \sqrt{n\alpha_n\beta_{\max}}$ and the first term has the desired rate.

To bound the spectral norm of $(D-D_P)P$ with high probability, we re-use the "light and heavy pairs" strategy of the previous proof. Define $\delta_i = d_i - \bar{d}_i$. We adopt the definitions of the previous section. Define $Q = (D-D_P)P$, we use again the fact that

$$\|(D-D_P)P\| \leqslant 4 \sup_{x,y\in T} x^\top Qy,$$

where T is the same grid. We decompose

$$x^\top Q y = \sum_{i,j \in \mathcal{L}(x,y)} x_i y_j p_{ij} \delta_i + \sum_{i,j \in \mathcal{H}(x,y)} x_i y_j p_{ij} \delta_i.$$

Recall that we have $|\delta_i| \leqslant d_{\max}$. In the proof of Theorem 3 we proved that $\sum_{i,j \in \mathcal{H}(x,y)} x_i y_j p_{ij} \leqslant \sqrt{n \alpha_n \beta_{\max}}$, and therefore with the same probability

$$\sum_{i,j \in \mathcal{H}(x,y)} x_i y_j p_{ij} \delta_i \lesssim d_{\max} \sqrt{n \alpha_n \beta_{\max}},$$

which is the desired complexity.

We must now handle the light pairs. We write $\delta_i = \sum_{\ell=1}^n \sum_{k=0}^t \beta_k w_{i\ell}^{(t-k)}$, and therefore

$$\sum_{i,j \in \mathcal{L}(x,y)} x_i y_j p_{ij} \delta_i = \sum_{i<\ell} \sum_k u_{i\ell k} w_{i\ell}^{(t-k)},$$

where

$$u_{i\ell k} = \beta_k \sum_j \left(x_i y_j p_{ij} 1_{(i,j) \in \mathcal{L}(x,y)} + x_\ell y_j p_{\ell j} 1_{(\ell,j) \in \mathcal{L}(x,y)} \right).$$

We want to apply Bernstein inequality. The random variables $w_{i\ell}^{(t-k)}$ are independent centered Bernoulli variables of parameters $p_{i\ell}^{(t-k)}$. By definition of light pairs we have

$$|u_{ijk}| \leqslant 2\beta_{\max} \alpha_n \bar{n}_{\max} \sqrt{\frac{\alpha_n}{n\beta_{\max}}} = 2\sqrt{\beta_{\max}} \alpha_n^{\frac{3}{2}} \frac{\bar{n}_{\max}}{\sqrt{n}},$$

since $\sum_j p_{ij}^{(t-k)} \leqslant \alpha_n \bar{n}_{\max}$. Then, using $(a+b)^2 \leqslant 2(a^2+b^2)$,

$$\sum_{i\ell k} Var(u_{i\ell k} w_{i\ell}^{(t-k)}) \leqslant \sum_k \beta_k^2 \left(\sum_{i\ell} p_{i\ell}^{(t-k)} \left(x_i \sum_j y_j p_{ij} + x_\ell \sum_j y_j p_{\ell j} \right)^2 \right)$$

$$\leqslant 2 \sum_k \beta_{\max}^2 \left(\sum_{i\ell} p_{i\ell}^{(t-k)} x_i^2 \left(\sum_j y_j p_{ij} \right)^2 + p_{i\ell}^{(t-k)} x_\ell^2 \left(\sum_j y_j p_{\ell j} \right)^2 \right)$$

$$\leqslant 2 C_\beta \beta_{\max} \bar{n}_{\max}^2 \alpha_n^3,$$

where we have used

$$\sum_j y_j p_{\ell j} = \sum_k \beta_k \sum_j y_j p_{\ell j}^{(t-k)} \leqslant \sum_k \beta_k \|y\| \alpha_n \sqrt{n_{\max} + \tau^2 n} \leqslant \alpha_n \sqrt{\bar{n}_{\max}},$$

and $\sum_{i\ell} p_{i\ell}^{(t-k)} x_i^2 \leqslant \alpha_n \bar{n}_{\max} \|x\|$. Hence, using Bernstein's inequality,

$$\mathbb{P}\left(|\sum_{i,j \in \mathcal{L}(x,y)} x_i y_j p_{ij} \delta_i| \geqslant t \right) \leqslant 2 \exp\left(-\frac{t^2/2}{2 C_\beta \beta_{\max} \bar{n}_{\max}^2 \alpha_n^3 + \frac{2}{3} \sqrt{\beta_{\max}} \alpha_n^{\frac{3}{2}} \frac{\bar{n}_{\max}}{\sqrt{n}} t} \right)$$

$$\mathbb{P}\left(|\sum_{i,j\in\mathcal{L}(x,y)} x_i y_j p_{ij}\delta_i| \geqslant c\bar{n}_{\max}\sqrt{n\beta_{\max}}\alpha_n^{\frac{3}{2}}\right) \leqslant 2\exp\left(-\frac{c^2/2}{2C_\beta + \frac{2c}{3}}\cdot n\right).$$

Using a union bound over T we can conclude.

A.4. Additional proofs

Proof of Lemma 5. Recall that we defined $\bar{\varepsilon} \stackrel{\text{def.}}{=} \max(\varepsilon_n, \sqrt{\log n/n})$. For any k, we have $p_{ij}^{(t-k)} = p_{ij}^{(t)}$ if $z_i^{t-k} = z_i^t$ and $z_j^{t-k} = z_j^t$, that is, if the nodes have not changed communities. Thus we write

$$\|P_{t-k} - P_t\|_F^2 = \sum_{i,j}(p_{ij}^{(t-k)} - p_{ij}^{(t)})^2 \left(1 - \mathbf{1}_{\{z_i^{t-k}=z_i^t\}}\mathbf{1}_{\{z_j^{t-k}=z_j^t\}}\right)$$

$$\leqslant \alpha_n^2 n^2 \left(1 - \left(\frac{1}{n}\sum_i \mathbf{1}_{\{z_i^{t-k}=z_i^t\}}\right)^2\right)$$

$$\alpha_n^2 n^2 \left(1 - \left(\frac{1}{n}\sum_i \mathbf{1}_{A_i}\right)^2\right)$$

where A_i is the event that $z_i^{t-k} = z_i^{t-k+1} = \ldots = z_i^t$. The A_i are independent and occur with probability $(1-\varepsilon_n)^k$. By Hoeffding inequality, for some $\delta > 0$ that we will fix later, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_i \mathbf{1}_{A_i} \leqslant (1-\varepsilon_n)^k - \delta\right) \leqslant 2e^{-2\delta^2 n}.$$

Therefore with probability at least $1 - 2e^{-2\delta^2 n}$,

$$\|P_{t-k} - P_t\|_F^2 \leqslant \alpha_n^2 n^2 \left(1 - \left((1-\varepsilon_n)^k - \delta\right)^2\right) \leqslant 2\alpha_n^2 n^2 \left(1 - (1-\varepsilon_n)^k + \delta\right).$$

Using $1 - x^k = (1-x)(1 + x + \ldots + x^{k-1}) \leqslant (1-x)k$ for $|x| \leqslant 1$ and $\varepsilon_n \leqslant \bar{\varepsilon}_n$ we get

$$\|P_{t-k} - P_t\|_F^2 \leqslant 2\alpha_n^2 n^2 \left(\min(1, k\bar{\varepsilon}_n) + \delta\right).$$

Then we choose $\delta \sim \bar{\varepsilon}_n \leqslant \min(1, k\bar{\varepsilon}_n)$, and use a union bound for $k = 1$ to $k \sim 1/\bar{\varepsilon}_n$ (since beyond that we have $\min(1, k\bar{\varepsilon}_n) = 1$) to conclude. $\qquad\square$

Proof of Lemma 7. Denote by $D = D(P_t)$, $\bar{D} = D(P_t^{\text{smooth}})$ the degree matrices of P_t and P_t^{smooth}, with d_i and $\bar{d}_i$ their elements. By assumption, we have $d_i, \bar{d}_i \geqslant d_{\min} \stackrel{\text{def.}}{=} \bar{n}_{\min}\alpha_n$ for all i. Therefore, by applying Lemma 11 we have

$$\|L(P_t) - L(P_t^{\text{smooth}})\| \leqslant \frac{\|P_t - P_t^{\text{smooth}}\|}{d_{\min}} + \frac{\|(D - \bar{D})P_t^{\text{smooth}}\|}{d_{\min}^2}.$$

From Lemma 4, we have $\|P_t - P_t^{\mathrm{smooth}}\| \lesssim C'_\beta \alpha_n \sqrt{\frac{n\bar{n}_{\max}\varepsilon_n}{\beta_{\max}}}$.

For the second term, we define D_{t-k} the diagonal degree matrix associated to P_{t-k}, such that

$$\|(D - \bar{D})P_t^{\mathrm{smooth}}\| \leqslant \sum_{k=0}^{t} \beta_k \|(D - D_{t-k})P_t^{\mathrm{smooth}}\|_F.$$

Denoting by $\bar{p}_{ij}$ the elements of P_t^{smooth}, recall that $\sum_j \bar{p}_{ij}^2 \leqslant \alpha_n^2 \bar{n}_{\max}$, and using $(a+b)^2 \leqslant 2(a^2+b^2)$ we have

$$\|(D - D_{t-k})P_t^{\mathrm{smooth}}\|_F^2 = \sum_i \left(\sum_\ell p_{i\ell}^{(t)} - p_{i\ell}^{(t-k)} \right)^2 \left(\sum_j \bar{p}_{ij}^2 \right)$$

$$\leqslant \alpha_n^2 \bar{n}_{\max} \sum_i \left(\sum_\ell p_{i\ell}^{(t)} - p_{i\ell}^{(t-k)} \right)^2$$

$$\leqslant \alpha_n^2 \bar{n}_{\max} \sum_i \left(\sum_\ell \left(\sqrt{p_{i\ell}^{(t)}} + \sqrt{p_{i\ell}^{(t-k)}} \right)^2 \right) \left(\sum_\ell \left(\sqrt{p_{i\ell}^{(t)}} - \sqrt{p_{i\ell}^{(t-k)}} \right)^2 \right)$$

$$\leqslant 2\alpha_n^2 \bar{n}_{\max} \sum_i \left(\sum_\ell p_{i\ell}^{(t)} + p_{i\ell}^{(t-k)} \right) \left(\sum_\ell \left(\sqrt{p_{i\ell}^{(t)}} - \sqrt{p_{i\ell}^{(t-k)}} \right)^2 \right)$$

$$\leqslant 4\alpha_n^3 \bar{n}_{\max}^2 \|P_t^{\odot \frac{1}{2}} - P_{t-k}^{\odot \frac{1}{2}}\|_F^2,$$

where $A^{\odot \frac{1}{2}}$ indicates element-wise square root. Repeating the proof of Lemma 4, for two SBM connection probability matrices P and P' between which only the nodes belonging to a set $\mathcal{S}$ have change community, we have

$$\|P^{\odot \frac{1}{2}} - (P')^{\odot \frac{1}{2}}\|_F^2 = \sum_{i \in \mathcal{S}} \sum_j \left(\sqrt{p_{ij}} - \sqrt{p'_{ij}} \right)^2 + \left(\sqrt{p_{ji}} - \sqrt{p'_{ji}} \right)^2$$

$$\leqslant 4 \sum_{i \in \mathcal{S}} \sum_j p_{ij} + p'_{ij} \leqslant 8|\mathcal{S}|\alpha_n \bar{n}_{\max}.$$

Therefore, $\|P_t^{\odot \frac{1}{2}} - P_{t-k}^{\odot \frac{1}{2}}\|_F^2 \lesssim \alpha_n \bar{n}_{\max} n \min(1, k\varepsilon_n)$, and $\|(D - D_{t-k})P_t^{\mathrm{smooth}}\|_F \leqslant \alpha_n^2 \bar{n}_{\max}^{\frac{3}{2}} \sqrt{n} \min(1, \sqrt{k\varepsilon_n})$. We conclude using the hypothesis on $\sum_k \beta_k \min(1, \sqrt{k\varepsilon_n})$. $\qquad \square$

A.5. Technical Lemma

Lemma 11. *Let $A, P \in \mathbb{R}^{n \times n}$ be symmetric matrices containing non-negative elements, assume that $d_i = \sum_j a_{ij}$ and $d_i^P = \sum_j p_{ij}$ are strictly positive, define $D = \mathrm{diag}(d_i)$, $D_P = \mathrm{diag}(d_i^P)$, $d_{\min} = \min_i(d_i, d_i^P)$. Then,*

$$\|L(A) - L(P)\| \leqslant \frac{\|A - P\|}{d_{\min}} + \frac{\|(D - D_P)P\|}{d_{\min}^2}.$$

Proof. Recall that $\|M\| = \sup_{x,y \in \mathcal{S}} x^\top M y$, where $\mathcal{S}$ is the Euclidean unit ball. Denote $Q = (D^{-\frac{1}{2}} - D_P^{-\frac{1}{2}})P$.

We write

$$
\begin{aligned}
L(A) - L(P) &= D^{-\frac{1}{2}} A D^{-\frac{1}{2}} - D_P^{-\frac{1}{2}} P D_P^{-\frac{1}{2}} \\
&= D^{-\frac{1}{2}}(A - P)D^{-\frac{1}{2}} + D^{-\frac{1}{2}} P D^{-\frac{1}{2}} - D_P^{-\frac{1}{2}} P D_P^{-\frac{1}{2}} \\
&= D^{-\frac{1}{2}}(A - P)D^{-\frac{1}{2}} + Q D^{-\frac{1}{2}} - D_P^{-\frac{1}{2}} Q^\top,
\end{aligned}
$$

and thus

$$
\|L(A) - L(P)\| \leqslant \frac{\|A - P\|}{d_{\min}} + \frac{2\|Q\|}{\sqrt{d_{\min}}}.
$$

Then, for $x, y \in \mathcal{S}$, we have

$$
\begin{aligned}
x^\top Q y &= \sum_{ij} x_i y_j p_{ij} \left(\frac{1}{\sqrt{d_i}} - \frac{1}{\sqrt{d_i^P}} \right) = \sum_{ij} x_i y_j p_{ij} \left(\frac{d_i^P - d_i}{\sqrt{d_i d_i^P}(\sqrt{d_i} + \sqrt{d_i^P})} \right) \\
&= \sum_{ij} x_i y_j p_{ij} \delta_i (d_i^P - d_i),
\end{aligned}
$$

where $\delta_i \overset{\text{def.}}{=} \frac{1}{\sqrt{d_i d_i^P}(\sqrt{d_i} + \sqrt{d_i^P})} \leqslant \frac{1}{2} d_{\min}^{-\frac{3}{2}}$. Since the p_{ij} are non-negative, the maximum over $x, y \in \mathcal{S}$ is necessarily reached when every term in the sum is non-negative, by choosing $y_j \geqslant 0$ and x_i with the same sign as $d_i^P - d_i$. Hence, $\sup_{x,y \in \mathcal{S}} \sum_{ij} x_i y_j p_{ij}(d_i^P - d_i) = \sup_{x,y \in \mathcal{S}} \sum_{ij} |x_i y_j p_{ij}(d_i^P - d_i)|$. Using this property,

$$
\begin{aligned}
\sup_{x,y \in \mathcal{S}} \sum_{ij} x_i y_j p_{ij} \delta_i (d_i^P - d_i) &\leqslant \sup_{x,y \in \mathcal{S}} \sum_{ij} |x_i y_j p_{ij} \delta_i (d_i^P - d_i)| \\
&\leqslant \frac{1}{2} d_{\min}^{-\frac{3}{2}} \sup_{x,y \in \mathcal{S}} \sum_{ij} |x_i y_j p_{ij}(d_i^P - d_i)| \\
&= \frac{1}{2} d_{\min}^{-\frac{3}{2}} \sup_{x,y \in \mathcal{S}} \sum_{ij} x_i y_j p_{ij}(d_i^P - d_i) \\
&= \frac{1}{2} d_{\min}^{-\frac{3}{2}} \|(D - D_P)P\|,
\end{aligned}
$$

which concludes the proof. $\qquad\square$

Acknowledgments

References

[1] ABBE, E. (2018). Community detection and stochastic block models: recent developments. *J. Mach. Learn. Res.* 1–86. MR3827065

[2] BACH, F. and JORDAN, M. (2006). Learning spectral clustering, with application to speech separation. *J. Mach. Learn. Res.* **7** 1963–2001. MR2274430

[3] BANDEIRA, A. S. and VAN HANDEL, R. (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.* **44** 2479–2506. MR3531673

[4] BHATTACHARYYA, S. and CHATTERJEE, S. (2018). Spectral clustering for multiple sparse networks: I. *arXiv:1805.10594* 1–24.

[5] BRUNA, J. and LI, X. (2017). Community Detection with Graph Neural Networks. 1–15. arXiv:1705.08415v2.

[6] CELISSE, A., DAUDIN, J. J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. MR2988467

[7] CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. *JMLR* **23** 1–23.

[8] CHI, Y., SONG, X., ZHOU, D., HINO, K. and TSENG, B. L. (2007). Evolutionary spectral clustering by incorporating temporal smoothness. In *KDD* 153–162.

[9] CHI, Y., SONG, X., ZHOU, D., HINO, K. and TSENG, B. L. (2009). On evolutionary spectral clustering. *ACM T. Knowl. Discov. D.* **3** 1–30.

[10] CHUNG, F. R. K. (1997). *Spectral graph theory* **92**. American Mathematical Soc. MR1421568

[11] COUILLET, R. and BENAYCH-GEORGES, F. (2016). Kernel spectral clustering of large dimensional data. *Electron. J. Stat.* **10** 1393–1454. MR3507369

[12] DHANJAL, C., GAUDEL, R. and CLÉMENÇON, S. (2014). Efficient eigen-updating for spectral graph clustering. *Neurocomputing* **131** 440–452.

[13] DIAO, P., GUILLOT, D., KHARE, A. and RAJARATNAM, B. (2016). Model-free consistency of graph partitioning. arXiv:1608.03860.

[14] GARCÍA TRILLOS, N. and SLEPČEV, D. (2018). A variational approach to the consistency of spectral clustering. *Appl. Comput. Harmon. Anal.* **45** 239–281. MR3824112

[15] GHASEMIAN, A., ZHANG, P., CLAUSET, A., MOORE, C. and PEEL, L. (2016). Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Phys. Rev. X* **6** 1–9.

[16] GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2009). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

[17] HAN, Q., XU, K. S. and AIROLDI, E. M. (2015). Consistent estimation of dynamic and multi-layer block models. In *ICML* **2** 1511–1520.

[18] HO, Q., SONG, L. and XING, E. P. (2011). Evolving cluster mixed-membership blockmodel for time-varying networks. *J. Mach. Learn. Res.* **15** 342–350.

[19] HOLLAND, P. W. (1983). Stochastic blockmodels: First Steps. *Soc. Networks* **5** 109–137. MR0718088

[20] HOLME, P. (2015). Modern temporal network theory: a colloquium. *Eur. Phys. J. B* **88**.

[21] KRZAKALA, F., MOORE, C., MOSSEL, E., NEEMAN, J., SLY, A., ZDE-BOROVÁ, L. and ZHANG, P. (2013). Spectral redemption: clustering sparse networks. *Proc. Natl. Acad. Sci. U.S.A.* 1–11. MR3174850

[22] KUMAR, A., SABHARWAL, Y. and SEN, S. (2004). A Simple Linear Time $(1 + \varepsilon)$-Approximation Algorithm for k-Means Clustering in Any Dimensions. In *FOCS* 454–462.

[23] LE, C. M., LEVINA, E. and VERSHYNIN, R. (2017). Concentration and regularization of random graphs. *Random Struct. Algor.* **51** 538–561. MR3689343

[24] LE, C. M., LEVINA, E. and VERSHYNIN, R. (2018). Concentration of random graphs and application to community detection. In *Proc. Int. Cong. of Math.* **3** 2913–2928. MR3966517

[25] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Stat.* **43** 215–237. MR3285605

[26] LEVIE, R., BRONSTEIN, M. M. and KUTYNIOK, G. (2019). Transferability of Spectral Graph Convolutional Neural Networks. arXiv:1907.12972. MR4353051

[27] LIU, F., CHOI, D., XIE, L. and ROEDER, K. (2018). Global spectral clustering in dynamic networks. *Proc. Natl. Acad. Sci. U.S.A.* **115** 927–932. MR3763702

[28] LLOYD, S. P. (1982). Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **28** 129–137. MR0651807

[29] LONGEPIERRE, L. and MATIAS, C. (2019). Consistency of the maximum likelihood and variational estimators in a dynamic stochastic block model. *Electron. J. Stat.* **13** 4157–4223. MR4021264

[30] MATIAS, C. and MIELE, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Stat. Soc. B* **79** 1119–1141. MR3689311

[31] MOSSEL, E., NEEMAN, J. and SLY, A. (2017). A proof of the block model threshold conjecture. *Combinatorica* 1–44. MR3876880

[32] NG, A. Y., JORDAN, M. I. and WEISS, Y. (2001). On Spectral Clustering: Analysis and Algorithm. In *NIPS* 849–856.

[33] NING, H., XU, W., CHI, Y., GONG, Y. and HUANG, T. (2007). Incremental Spectral Clustering With Application to Monitoring of Evolving Blog Communities. 261–272.

[34] OLIVEIRA, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. arXiv:0911.0600.

[35] PAUL, S. and CHEN, Y. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *Ann. Stat.* **48** 230–250. MR4065160

[36] PENG, R., SUN, H. and ZANETTI, L. (2015). Partitioning Well-Clustered Graphs: Spectral Clustering Works! **40** 1–33. MR3629439

[37] Pensky, M. and Zhang, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electron. J. Stat.* **13** 678–709. MR3914178

[38] Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Stat.* **39** 1878–1915. MR2893856

[39] Sarkar, P. and Bickel, P. J. (2015). Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Stat.* **43** 962–990. MR3346694

[40] Tang, M. and Priebe, C. E. (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Ann. Stat.* **46** 2360–2415. MR3845021

[41] Tremblay, N. and Loukas, A. (2020). Approximating Spectral Clustering via Sampling: a Review. In *Sampling Techniques for Supervised or Unsupervised Tasks* 129–183. Springer.

[42] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* **17** 395–416. MR2409803

[43] Von Luxburg, U., Belkin, M. and Bousquet, O. (2008). Consistency of spectral clustering. *Ann. Stat.* **36** 555–586. MR2396807

[44] Wang, Y., Chakrabarti, A., Sivakoff, D. and Parthasarathy, S. (2017). Fast Change Point Detection on Dynamic Social Networks. In *IJCAI* 2992–2998.

[45] Xu, D. and Knight, J. (2010). Continuous Empirical Characteristic Function Estimation of Mixtures of Normal Parameters. *Econom. Rev.* **30** 25–50. MR2747367

[46] Xu, K. S. (2015). Stochastic Block Transition Models for Dynamic Networks. In *AISTAT* 1–23.

[47] Xu, K. S. and Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Sel. Topics Signal Process.* **8** 552–562.

[48] Xu, K. S., Kliger, M. and Iii, A. O. H. (2010). Evolutionary spectral clustering with adaptative forgetting factor. In *ICASSP* 2174–2177.

[49] Yang, T., Chi, Y., Zhu, S., Gong, Y. and Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks – A Bayesian approach. *Mach. Learn.* **82** 157–189. MR3108191