Proceedings of Machine Learning Research vol 99:1–30, 2019 32nd Annual Conference on Learning Theory

Robustness of spectral methods for community detection

Ludovic Stephan LUDOVIC.STEPHAN@INRIA.FR
Inria, École Normale Supérieure, PSL Research University

Laurent Massoulié LAURENT.MASSOULIE@INRIA.FR
Inria, École Normale Supérieure, PSL Research University, Microsoft Research-Inria Joint Centre

Editors: Alina Beygelzimer and Daniel Hsu

Abstract

The present work is concerned with community detection. Specifically, we consider a random graph drawn according to the stochastic block model: its vertex set is partitioned into blocks, or communities, and edges are placed randomly and independently of each other with probability depending only on the communities of their two endpoints. In this context, our aim is to recover the community labels better than by random guess, based only on the observation of the graph.

In the sparse case, where edge probabilities are in $O(1/n)$, we introduce a new spectral method based on the distance matrix $D^{(\ell)}$, where $D^{(\ell)}_{ij} = 1$ iff the graph distance between i and j, noted $d(i, j)$ is equal to ℓ. We show that when $\ell \sim c \log(n)$ for carefully chosen c, the eigenvectors associated to the largest eigenvalues of $D^{(\ell)}$ provide enough information to perform non-trivial community recovery with high probability, provided we are above the so-called Kesten-Stigum threshold. This yields an efficient algorithm for community detection, since computation of the matrix $D^{(\ell)}$ can be done in $O(n^{1+\kappa})$ operations for a small constant κ.

We then study the sensitivity of the eigendecomposition of $D^{(\ell)}$ when we allow an adversarial perturbation of the edges of G. We show that when the considered perturbation does not affect more than $O(n^{\varepsilon})$ vertices for some small $\varepsilon > 0$, the highest eigenvalues and their corresponding eigenvectors incur negligible perturbations, which allows us to still perform efficient recovery.

Our proposed spectral method therefore: i) is robust to larger perturbations than prior spectral methods, while semi-definite programming (or SDP) methods can tolerate yet larger perturbations; ii) achieves non-trivial detection down to the KS threshold, which is conjectured to be optimal and is beyond reach of existing SDP approaches; iii) is faster than SDP approaches.

1. Introduction

1.1. Background

Community detection is the task of finding large groups of similar items inside a large relationship graph, where it is expected that related items are (in the assortative case) more likely to be linked together. The Stochastic Block Model (abbreviated in SBM) has been designed by Holland et al. (1983) to analyze the performance of algorithms for this task; it consists in a random graph G whose edge probabilities depend only on the community membership of their endpoints. Since then, a large number of articles have been devoted to the study of this model; a survey of these results can be found in Abbe (2017), or in Fortunato (2010) for a more general view on community detection.

The sparse case, when edge probabilities are in $O(1/n)$, is known to be much harder to study than denser models; the existence of a positive portion of isolated vertices makes complete reconstruction impossible, and studies usually focus on partial recovery of the community structure. Insights on this topic often stem from statistical physics; in the two-community case, Decelle et al. (2011) conjectured the existence of a threshold for reconstruction, as well as its exact value; this conjecture was then proved in Mossel et al. (2015) for the first part, Massoulié (2014) and Mossel et al. (2013) for the converse part. Similarly, in the general case, a method was first presented in Krzakala et al. (2013) and then proven to work in Bordenave et al. (2015) – bar a technical condition – and Abbe and Sandon (2016).

The main issue in the sparse setting is that the usual method relying on the eigenvectors of the adjacency matrix of G fails due to the lack of separation of the eigenvalues. Consequently, a wide array of alternative spectral methods have been designed, relying on the spectrum of a matrix associated to G. More precisely, the eigenvectors associated to the highest eigenvalues of those matrices will often carry some information about the community structure of G, enough for partial reconstruction. Examples include the path expansion matrix used in Massoulié (2014), or the non-backtracking matrix in Krzakala et al. (2013).

Additionally, other types of methods can be used in this setting: for example, the semi-definite programming (or SDP) algorithm relaxes the problem into a convex optimization one, which can then be approximately solved (see for example Montanari and Sen (2016)).

An important feature of real-life networks that is missing from the SBM is the existence of small-scale regions of higher density, that arise from phenomena unrelated to the community structure. For this reason, a common variant of the SBM is the addition of small cliques to the generated random graph. Commonly-used spectral methods, for example those relying on the non-backtracking matrix in Bordenave et al. (2015), are known to fail in this setting, due to the apparition of localized eigenvectors, with no ties to the community structure, and corresponding to large eigenvalues – see Zhang (2016) for a comparison of those methods, as well as a proposed heuristic to deal with those localized vectors by lowering their associated eigenvalues. SDP methods are the most studied for this problem, due to their natural stability; in particular, Makarychev et al. (2016) show a reconstruction algorithm that is robust to the adversarial addition of $o(n)$ edges, in the case of an arbitrary number of communities; this was also shown independently by Moitra et al. (2016). However, all the SDP methods mentioned here fail to reach the KS threshold by at least a large constant, with only Montanari and Sen (2016) approaching it as the average degree increases.

After completing this work we became aware of the article of Abbe et al. (2018). It establishes results akin to ours on robustness (although with a different definition thereof) of spectral methods for detection in SBM. We use however a slightly different matrix, and our results apply to an arbitrary number of blocks, whereas they only consider SBM with two blocks.

1.2. Summary of main results

This article focuses on the Stochastic Block Model, as defined in Holland et al. (1983); we recall here a succinct definition:

Definition 1 *Let $r \in \mathbb{N}$ be fixed, W be a $r \times r$ symmetric matrix with nonnegative entries, and π a probability vector on $[r]$. A random graph $G = (V, E)$ with $|V| = n$ is said to be distributed according to the Stochastic Block Model (or SBM) with r blocks and parameters (W, π) if:*

(i) each vertex $v \in V$ is assigned a type *$\sigma(v)$ sampled independently from π,*

(ii) any two vertices u, v in V are joined with an edge randomly and independently from every other edge, with probability

$$\min(\frac{W_{\sigma(u),\sigma(v)}}{n}, 1).$$

Given a random graph G sampled according to the above model (with the types of each vertex hidden), the aim is to estimate the type assignment σ from the observation of G only. However, since there is a positive proportion of isolated vertices, perfect reconstruction is theoretically impossible; we will thus only focus on retrieving only a positive proportion of the types. Our metric for reconstruction will be the following:

Definition 2 *Let σ be the true type estimation, and $\hat{\sigma}$ a type estimate of σ; the* empirical overlap *between σ and $\hat{\sigma}$ is defined as:*

$$\mathrm{ov}(\sigma, \hat{\sigma}) = \max_{\tau \in \mathfrak{S}_r} \frac{1}{n} \sum_{v=1}^{n} \mathbf{1}_{\hat{\sigma}(v) = \tau(\sigma(v))} - \max_{k \in [r]} \pi_k, \tag{1}$$

where $\mathfrak{S}_r$ is the set of permutations of $[r]$. For a given algorithm leading to estimates $\hat{\sigma}$ for all n, we say that this algorithm achieves partial reconstruction if

$$\liminf_{n \to \infty} \ \mathrm{ov}(\sigma, \hat{\sigma}) > 0 \quad \textit{w.h.p.} \tag{2}$$

Spectral methods in denser settings (with average degrees of about $\log(n)$) usually consist in clustering the eigenvectors of the adjacency matrix of G; however those methods are known to fail in sparser graphs (see Abbe (2017). As a result, different (and more complex) matrices are needed:

Definition 3 *Let G be any graph, and ℓ be a positive integer. We define two matrices associated with G:*

(i) the path expansion matrix *$B^{(\ell)}$ (studied in Massoulié (2014)), whose (i, j) coefficient counts the number of self-avoiding paths (that is, paths that do not go through the same vertex twice) of length ℓ between i and j,*

(ii) the distance matrix *$D^{(\ell)}$, defined by $D^{(\ell)}_{ij} = 1$ if $d(i, j) = \ell$ and 0 otherwise, where d denotes the usual graph distance.*

We are now ready to state our first result:

Theorem 4 *Assume that $\pi \equiv 1/r$, W is a stochastic positive regular matrix, and that the two highest eigenvalues μ_1, μ_2 of W satisfy the condition:*

$$\mu_2^2 > \mu_1.$$

Then there exists an algorithm, based only on an eigenvector of $B^{(\ell)}$ associated with its second highest eigenvalue, that achieves partial reconstruction whenever $\ell \sim \delta \log(n)$ for small enough δ.

The same algorithm also achieves partial reconstruction when applied to $D^{(\ell)}$ instead of $B^{(\ell)}$, with the same conditions on ℓ.

Regardless of the change of matrices, this is already an improvement on Bordenave et al. (2015); indeed, we managed to remove a technical asymmetry condition on W (namely, the existence of a simple eigenvector associated with a high eigenvalue).

We now move on to study the stability of our algorithm; as opposed to most papers that classify the difficulty of an adversary according to the number of altered edges, ours considers the number of affected vertices.

Definition 5 *Let $\gamma := \gamma(n)$ be a positive integer, and G any graph on n vertices. The adversary of* strength γ *is allowed to arbitrarily add and remove edges at will, as long as the number of vertices affected (i.e. vertices that are endpoints of altered edges) is at most γ.*

Our main result on stability is then the following:

Theorem 6 *Under the same assumptions as Theorem 4, let G a graph generated via SBM, and $\tilde{G}_\gamma$ a graph obtained when perturbed by an adversary of strength γ.*
Then, assuming

$$\gamma = o\left(\frac{(\mu_2^2/\mu_1)^{\ell/2}}{\log(n)}\right),$$

the algorithm of Theorem 4 still achieves partial reconstruction on $\tilde{G}_\gamma$.

The above result on γ is the best possible, up to a factor of $\log(n)$.

Compared to the spectral algorithm in Bordenave et al. (2015), this is a substantial improvement: their algorithm is known (see e.g. Zhang (2016)) to be highly unstable w.r.t edge addition. In contrast, the above result reaches a perturbation of size a small power of n (since ℓ is of order $\log(n)$). This is sharp, and thus still far from the $o(n)$ bound achieved by various SDP methods (notably Makarychev et al. (2016); Montanari and Sen (2016)); this discrepancy is likely a result of delicate graph properties involved in spectral algorithms that make them more sensitive to perturbations.

However, our result still has several advantages compared to the other cited methods, namely:

(i) the threshold for partial reconstruction in our method is exactly the KS threshold, whereas SDP-based methods require a slightly stronger condition, especially when the mean degree of G is low.

(ii) as will be proved later, the running time of our algorithm is at most $O(n^{13/12})$, which is much faster that the usual methods for SDP algorithms.

(iii) Finally, all the SDP methods mentioned throughout this paper only consider the symmetric case even in the case of multiple communities.

1.3. Detailed setting and results

STOCHASTIC BLOCK MODEL

Consider the SBM as defined above; following Bordenave et al. (2015), we introduce $\Pi = \mathrm{diag}(\pi_1, \dots, \pi_r)$ and define the *mean progeny matrix* $M = \Pi W$; the eigenvalues of M are the same as those of the symmetric matrix $S = \Pi^{1/2} W \Pi^{1/2}$ and in particular are real. We denote them by

$$\mu_1 \geq |\mu_2| \geq \dots \geq |\mu_r|.$$

We shall make the following regularity assumptions: first,

$$\mu_1 > 1 \quad \text{and } M \text{ is positive regular,} \tag{3}$$

i.e. the coefficients of M^t are all positive for some t. Secondly, each type of vertex has the same asymptotic average degree, that is

$$\sum_{i=1}^{r} M_{ij} = \sum_{i=1}^{r} \pi_i W_{ij} = \alpha \quad \text{for all } j \in [r]. \tag{4}$$

In this case, the matrix $M^* = M/\alpha$ is a stochastic matrix and therefore

$$\mu_1 = \alpha > 1. \tag{5}$$

Since $M = \Pi^{-1/2} S \Pi^{1/2}$, M is diagonalizable; let $(\phi_1, \dots, \phi_r)$ be a basis of normed left eigenvectors for M, that is

$$\phi_i^\top M = \mu_i \phi_i^\top \quad \text{for all } i \in [r]. \tag{6}$$

Condition (4) implies that $\phi_1 = \mathbf{1}/\sqrt{r}$, where $\mathbf{1}$ is the all-ones vector.

It has been shown in Bordenave et al. (2015) and Abbe and Sandon (2016) that polynomial-time algorithms achieve partial reconstruction when the following condition, called the Kesten-Stigum threshold, is verified:

$$\tau := \mu_2^2/\mu_1 > 1. \tag{7}$$

The quantity τ is commonly referred to as the *signal-to-noise ratio*.

Alternatively, we define r_0 such that

$$\mu_{r_0+1}^2 \leq \mu_1 < \mu_{r_0}^2. \tag{8}$$

Therefore, the condition (7) is equivalent to $r_0 > 1$.

In the two-community case, the above condition is equivalent to the possibility of reconstruction (see Massoulié (2014), Mossel et al. (2015)). However, in the general setting ($r > 4$), non-polynomial algorithms can achieve partial reconstruction even below this threshold. This was originally conjectured in Decelle et al. (2011), and more recently proven in Abbe and Sandon (2016).

PATH EXPANSION MATRIX

In Massoulié (2014), an algorithm for partial reconstruction in the two-community case makes use of the path expansion matrix $B^{(\ell)}$. Our first aim is to extend the result from this paper to the general case; we first define for all $k \in [r]$ the vectors χ_k and φ_k by

$$\chi_k(v) = \phi_k(\sigma(v)) \quad \text{and} \quad \varphi_k = \frac{B^{(\ell)}\chi_k}{\|B^{(\ell)}\chi_k\|}. \tag{9}$$

Let $\lambda_1(B^{(\ell)}) \geq |\lambda_2(B^{(\ell)})| \geq |\lambda_n(B^{(\ell)})|$ be the eigenvalues of $B^{(\ell)}$ ordered by absolute value; our first theorem is an extension of Theorem 2.1 in Massoulié (2014):

Theorem 7 *Consider a graph G generated as above, and let $\ell \sim \kappa \log_\alpha(n)$, with $\kappa < 1/12$. Then, with probability going to 1 as n goes to $+\infty$:*

(i) $\lambda_k(B^{(\ell)}) = \Theta(\mu_k^\ell)$ *for* $k \in [r_0]$,

(ii) For $k > r_0$, $\lambda_k(B^{(\ell)}) = O(\log(n)^c\, \alpha^{\ell/2})$ *for some constant* $c > 0$.

*Furthermore, consider μ such that $\mu^2 > \alpha$ and μ is an eigenvalue of multiplicity d of **M**. Let $\phi^{(1)}, \dots, \phi^{(d)}$ be an orthonormal basis of eigenvectors of M associated to μ, and $\varphi^{(1)}, \dots, \varphi^{(d)}$ the vectors defined as in (9).*
There exist orthogonal vectors $\xi^{(1)}, \dots, \xi^{(d)}$ in $\mathbb{R}^n$ such that the following holds:

(i) for all i, $\xi^{(i)}$ is an eigenvector of $B^{(\ell)}$, with associated eigenvalue $\Theta(\mu^\ell)$

(ii) there exists an orthogonal matrix $Q \in O(d)$ such that

$$\|\boldsymbol{\varphi} Q - \boldsymbol{\xi}\|_2 = O\left(\alpha^{\ell/2}\mu^{-\ell}\right),$$

where $\boldsymbol{\varphi}$ (resp. $\boldsymbol{\xi}$) is the $n \times d$ matrix whose columns are the $\varphi^{(i)}$ (resp. the $\xi^{(i)}$).

The above theorem does not yield immediately an algorithm for community reconstruction; however, adapting the one found in Bordenave et al. (2015), we designed the following:

Algorithm 8 *Let ξ be an eigenvector of $B^{(\ell)}$ associated to the eigenvalue $\lambda_2(B^{(\ell)})$, normalized such that $\|\xi\|^2 = n$, and K an arbitrary large constant. First, partition V in two sets (I^+, I^-) via the following procedure: put v in I^+ with probability*

$$\mathbb{P}(v \in I^+ \mid \xi) = \frac{1}{2} + \frac{1}{2K}\,\xi(v)\,\mathbf{1}_{|\xi(v)| \leq K}$$

Then, assign the label 1 to every vertex in I^+ and label 2 to every vertex in I^-.

We then have the following theorem:

Theorem 9 *Assume that $\pi \equiv 1/r$, and that $r_0 > 1$, i.e. that we are above the Kesten-Stigum threshold. Then Algorithm 8 yields an asymptotically positive overlap when $n \to \infty$ for some choice of K.*

Note that we don't need the asymmetry condition from Bordenave et al. (2015) anymore; our algorithm can deal with multiple eigenvalues as well. Additionally, an explicit value for K is derived in the appendix, which makes our algorithm easy to implement and eliminates the need for "magic" constants, such as the ones in Zhang (2016) or Abbe and Sandon (2016).

A crucial feature of this algorithm is that it depends only on the second eigenvalue of $B^{(\ell)}$; for any perturbation that leaves the r_0 highest eigenvalues – or even the second highest – unchanged, the result from Theorem 9 will hold.

THE DISTANCE MATRIX

We introduce now the *distance matrix* $D^{(\ell)}$, defined by $D^{(\ell)}_{ij} = 1$ if and only if $d(i, j) = \ell$, where d is the distance in G. This matrix, while sparser than $B^{(\ell)}$, retains much of the desired spectral properties. In particular, we have the following theorem:

Theorem 10 *Assume that condition (7) holds, and set ℓ such that $\ell \sim \kappa \log_\alpha(n)$, where κ is a constant such that $\kappa < 1/12$. Then the results of Theorem 7 hold with the matrix $B^{(\ell)}$ replaced by $D^{(\ell)}$.*

As a result, Algorithm 8 will still succeed when applied to the matrix $D^{(\ell)}$.

GRAPH PERTURBATION

As mentioned in the introduction, community detection algorithms have to be resilient to the presence of small cliques (or denser subgraphs) to be useful in practice, since this kind of pattern is often present in real-life networks. We chose to focus here on adversarial perturbations, as defined in the summary, whereas other papers (mainly Abbe et al. (2018)) focus instead on other random graph models, more prone to small loops and cliques.

As shown in Zhang (2016), the usual spectral methods do not fare well against adversarial (or even random) perturbation, especially when the added subgraph contains several cliques. This is especially the case for the non-backtracking matrix in Bordenave et al. (2015), but also the path expansion matrix in Massoulié (2014).

However, the distance matrix is more stable to clique addition, since it does not count the number of paths between two vertices – which is affected significantly by small perturbations. We can therefore allow a perturbation of size up to a small power of n, as stated in the following theorem:

Theorem 11 *Let G be an SBM as above, with $\pi_i \equiv 1/r$. Assume that $r_0 > 1$, and recall that $\tau = \mu_2^2/\mu_1 > 1$ is the signal-to-noise ratio.*

Then, if $\gamma = o(\tau^\ell / \log(n))$, then Algorithm 8 based on the distance matrix recovers the original communities with asymptotically positive overlap, even after a perturbation affecting at most γ vertices.

The controls in the above theorem can be shown to be sharp, up to a factor of $\log(n)$:

Theorem 12 *With the same assumptions as above, let $D^{(\ell)}$ be the distance matrix of G and $\tilde{D}^{(\ell)}$ the one of the graph after the adversarial perturbation.*

If $\gamma = \Omega(\tau^\ell)$, then there exists a perturbation of size at most γ such that $\tilde{D}^{(\ell)}$ has an eigenvalue of size $\Omega(\mu_2^\ell)$, with associated eigenvector asymptotically perpendicular to the first r_0 ones of $D^{(\ell)}$.

Therefore, we cannot guarantee the stability of the eigenvectors of $D^{(\ell)}$ when the perturbation affects too many vertices. This means that the best bound we can get on the size of allowed perturbations of the matrix $D^{(\ell)}$ is τ^ℓ, which we can rewrite as

$$\tau^\ell = n^{\kappa \log_\alpha(\tau)}.$$

The spectral method on the distance matrix is thus robust to perturbations of size at most n^ε, with $\varepsilon = \kappa \log_\alpha(\tau)$ going to zero as we approach the KS threshold.

1.4. Notations and outline of the paper

Throughout this paper, we will make use of the following notation: for two functions f, g, we say that $f = \widetilde{O}(g)$ if there exists a constant c such that $f = O\left(\log(n)^c \cdot g\right)$. We similarly define the notations $\widetilde{\Theta}$ and $\widetilde{\Omega}$.

The next Section is devoted to the study of the spectral structure of $B^{(\ell)}$; we also state there an important theorem on spectral perturbation that will be useful for the study of matrix $D^{(\ell)}$ as well. In Section 3, we study the distance matrix $D^{(\ell)}$ and introduce a method to deal with perturbations of this matrix. We then leverage this method to obtain bounds on the size of allowed perturbations.

2. Spectral structure of $B^{(\ell)}$

2.1. A theorem on eigenspace perturbation

In the following, we'll need a way to link the operator norm of a matrix perturbation to the consequent perturbation of its eigenvectors. This is provided by the following variant of the Davis-Kahan $\sin\theta$ theorem (Yu et al. (2015), Theorem 2):

Theorem 13 *Let $\Sigma, \hat{\Sigma}$ be symmetric $n \times n$ matrices, with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$ and $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_n$ respectively. Fix $1 \leq r \leq s \leq n$ and assume that $\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1}) > 0$, where we define $\lambda_0 = +\infty$ and $\lambda_{n+1} = -\infty$.*

Let $d = s - r + 1$, and let $V = (v_r, \ldots, v_s)$ and $\hat{V} = (\hat{v}_r, \ldots, \hat{v}_s)$ have orthonormal columns satisfying $\Sigma v_j = \lambda_j v_j$ and $\hat{\Sigma}\hat{v}_j = \hat{\lambda}_j \hat{v}_j$ for $j \in \{r, \ldots, s\}$.

Then there exists an orthogonal matrix $Q \in O(d)$ such that

$$\|VQ - \hat{V}\|_F \leq \frac{2\sqrt{2d}\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}}}{\min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1})}. \tag{10}$$

2.2. Strategy of proof

We present here the main ideas of the proof, and defer its full version to the appendix. The first step is an adaptation of Proposition 19 from Bordenave et al. (2015):

Proposition 14 *Let $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < 1/12$. Define, for $k \in [r]$,*

$$\theta_k = \|B^{(\ell)}\varphi_k\| \text{ and } \zeta_k = \frac{B^{(\ell)}\varphi_k}{\theta_k}, \tag{11}$$

with φ_k as in (9).

Then, with high probability, we have the following estimations for every $\gamma < 1/2$:

(i) $\theta_k = \Theta(\mu_k^\ell)$ *for* $k \in [r_0]$*,*

(ii) $|\langle \varphi_j, \varphi_k \rangle| = \widetilde{O}(\alpha^{3\ell/2} n^{-\gamma/2})$ *for* $j \neq k \in [r_0]$*,*

(iii) $|\langle \zeta_j, \varphi_k \rangle| = \widetilde{O}(\alpha^{2\ell} n^{-\gamma/2})$ *for* $j \neq k \in [r_0]$*.*

Now, let $(z_1, \dots, z_{r_0})$ be the Gram-Schmidt orthonormalization of $(\varphi_1, \dots, \varphi_{r_0})$, and define

$$D = \sum_{k=1}^{r_0} \theta_k z_k z_k^\top.$$

The non-zero eigenvalues of D are thus the θ_k, with corresponding eigenvectors z_k. Then, using the asymptotic orthogonality properties of Proposition 14, we prove the following:

Proposition 15 *For all* $k \in [r_0]$*,* z_k *is asymptotically parallel to* φ_k*.*
Furthermore,

$$\|B^{(\ell)} - D\|_{\mathrm{op}} = \widetilde{O}(\alpha^{\ell/2}). \tag{12}$$

Theorem 7 then results from a simple application of the Weyl inequality (Weyl (1912)) and Theorem 13:

Proof (of Theorem 7): let $\theta_k = 0$ for $k > r_0$; the eigenvalues of D are then exactly the θ_i for $i \leq n$.
By Weyl's inequality, we have for all $i \in [n]$

$$|\lambda_i(B^{(\ell)}) - \theta_i| = \widetilde{O}(\alpha^{\ell/2}).$$

Since $\theta_k = \Theta(\mu_k^\ell)$ for $k \in [r_0]$, this implies the statements (i) and (ii) of the Theorem.

We now define $z^{(1)}, \dots, z^{(d)}$ as the z_i associated to $\varphi^{(1)}, \dots, \varphi^{(d)}$, and $\mathbf{z}$ as in Theorem 7. Applying inequality (10) to $B^{(\ell)}$ and D yields the existence of an orthogonal matrix $Q \in O(d)$ such that

$$\|\mathbf{z}Q - \boldsymbol{\xi}\| = O\left(\alpha^{\ell/2}\mu^{-\ell}\right),$$

and the proof of Proposition 15 shows that $\|z^{(i)} - \varphi^{(i)}\| = O\left(\alpha^{\ell/2}\mu^{-\ell}\right)$ for all i. Using the triangular inequality (and the fact that Q preserves the norm) completes the proof of Theorem 7. ■

2.3. A new reconstruction algorithm

We now sketch the proof for Theorem 9; it hinges on one key lemma, whose proof (adapted from Bordenave et al. (2015)) is in the appendix:

Lemma 16 *Let* ξ *be as in Algorithm 8. For all* $i \in [r]$*, there exists a random variable* X_i *such that for every* $K > 0$ *that is a continuity point of* X_i*, in probability,*

$$\frac{1}{n} \sum_{v \in V} \mathbf{1}_{\sigma(v)=i}\, \xi(v) \mathbf{1}_{|\xi(v)| \leq K} \to \pi(i)\, \mathbb{E}[X_i \mathbf{1}_{|X_i| \leq K}],$$

where the convergence is independent from ξ*.*

Furthermore, we have

$$\sum_{i\in[r]} \mathbb{E}[X_i] = 0 \quad \text{and} \quad \sum_{i\in[r]} \mathbb{E}[X_i]^2 > c \tag{13}$$

for some absolute constant $c > 0$, and for all $\varepsilon > 0$ there exists a choice of M (independent from the chosen eigenvector ξ) such that

$$\left|\mathbb{E}[X_i \mathbf{1}_{|X_i|\leq K}] - \mathbb{E}[X_i]\right| < \varepsilon \tag{14}$$

In particular, Lemma 16 implies that there exist i and j such that $|\mathbb{E}[X_i] - \mathbb{E}[X_j]| > \sqrt{c}$.

We then use a concentration bound to show that for all i, in probability,

$$\frac{1}{n}\sum_{v\in V} \mathbf{1}_{\sigma(v)=i}\, \mathbf{1}_{v\in I^+} \to \pi(i)\left(\frac{\mathbb{E}[X_i \mathbf{1}_{|X_i|\leq K}]}{2K} + 1/2\right) := \pi(i)\tilde{p}_i \tag{15}$$

where the convergence is independent from ξ.

Assume now that $\pi \equiv 1/r$; from Lemma 16, for a large enough M there exists a $\delta > 0$ such that $\tilde{p}_i > \tilde{p}_j + \delta$. Assign label 1 to I^+ and 2 to I^-, and let τ be a permutation such that $\tau(i) = 1$ and $\tau(j) = 2$. The overlap achieved by τ is thus

$$\frac{1}{n}\sum_{v\in V} \mathbf{1}_{\sigma(v)=i}\, \mathbf{1}_{v\in I^+} + \frac{1}{n}\sum_{v\in V} \mathbf{1}_{\sigma(v)=j}\, \mathbf{1}_{v\in I^-} - \frac{1}{r} = \frac{1}{r}\left(\tilde{p}_i + 1 - \tilde{p}_j\right) - \frac{1}{r} > \frac{\delta}{r}, \tag{16}$$

which completes the proof of Theorem 9.

3. Study of the matrix $D^{(\ell)}$

3.1. From $B^{(\ell)}$ to $D^{(\ell)}$

The first aim of this section is to prove Theorem 10, i.e. that we can replace matrix $B^{(\ell)}$ by $D^{(\ell)}$ in the algorithm from Theorem 9. Directly proving this theorem is hard, because of the lack of a decomposition such as the one in Lemma 25 for $D^{(\ell)}$. However, in view of the proof of Theorem 7 above, it is sufficient to prove the following proposition:

Proposition 17 *Let G be a SBM as above, and $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < 1/12$. Let $B^{(\ell)}$ be the path expansion matrix of G, and $D^{(\ell)}$ its distance matrix. Then, with high probability:*

$$\rho(B^{(\ell)} - D^{(\ell)}) = \widetilde{O}(\alpha^{\ell/2}), \tag{17}$$

where ρ is the spectral radius of a matrix.

For ease of notation, let $\Delta^{(\ell)} = B^{(\ell)} - D^{(\ell)}$; we first notice that $\Delta^{(\ell)}$ is a $0-1$ matrix:

Lemma 18 *Let $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < 1/12$. For all vertices $i, j \in \{1, \dots, n\}$,*

$$0 \leq \Delta^{(\ell)}_{ij} \leq 1. \tag{18}$$

Furthermore, if $\Delta^{(\ell)}_{ij} = 1$, then there exists a cycle $\mathcal{C}$ such that:

$$d(i, \mathcal{C}) + d(j, \mathcal{C}) \leq \ell. \tag{19}$$

Define now a matrix $P^{(\ell)}$ by $P^{(\ell)}_{ij} = 1$ if there is a cycle $\mathcal{C}$ such that $d(i, \mathcal{C}) + d(j, \mathcal{C}) \leq \ell$. By the previous lemma, we have $\Delta^{(\ell)}_{ij} \leq P^{(\ell)}_{ij}$ for all (i, j), and the Perron-Frobenius theorem implies:

$$\rho(\Delta^{(\ell)}) \leq \rho(P^{(\ell)}). \tag{20}$$

It remains then to bound the spectral radius of $P^{(\ell)}$; the key lemma is the following:

Lemma 19 *For a given cycle $\mathcal{C}$, let $P^{(\ell)}_{\mathcal{C}}$ be the matrix defined by $P^{(\ell)}_{\mathcal{C},ij} = 1$ if $d(i, \mathcal{C}) + d(j, \mathcal{C}) \leq \ell$, and $V_{\mathcal{C}}$ the set of vertices such that $d(i, \mathcal{C}) \leq \ell$. Then:*

(i) $P^{(\ell)}_{\mathcal{C}}$ *is zero outside of* $V_{\mathcal{C}} \times V_{\mathcal{C}}$,

(ii) $\rho(P^{(\ell)}) = \max_{\mathcal{C}} \rho(P^{(\ell)}_{\mathcal{C}})$.

By part (ii) of the above lemma, it is sufficient to bound $\rho(P^{(\ell)}_{\mathcal{C}})$ for a given cycle $\mathcal{C}$ in $\mathcal{G}$; using part (i) of the above lemma, we can restrict our study to the subspace spanned by the vertices in $V_{\mathcal{C}}$.

Let v be a normed vector of size $|V_{\mathcal{C}}|$ corresponding to the highest eigenvalue of $P^{(\ell)}_{\mathcal{C}}$; as the coefficient (i, j) of $P^{(\ell)}_{\mathcal{C}}$ only depends on the distance of i and j to $\mathcal{C}$, we likewise group the coefficients of v by their distance t to $\mathcal{C}$, and write

$$v = (v_{tj})_{\substack{0 \leq t \leq \ell \\ 1 \leq j \leq S_t(\mathcal{C})}}.$$

We then have:

$$v^\top P_{\mathcal{C}} v = \sum_{t+u \leq \ell} \sum_{i,j} v_{ti}\, v_{uj} = \sum_{t+u \leq \ell} \left(\sum_i v_{ti} \right) \left(\sum_j v_{uj} \right).$$

By the Perron-Frobenius theorem, the coefficients of v are non-negative. For a given t, the coefficients v_{ti} are necessarily equal; otherwise, we could increase $\sum v_{ti}$ while leaving $\sum v_{ti}^2$ fixed, which leads to increasing $v^\top P^{(\ell)}_{\mathcal{C}} v$ while keeping $\|v\|^2$ constant: this contradicts the definition of v.

Writing $v_{ti} = v_t$ for all $1 \leq i \leq S_t(\mathcal{C})$; we get:

$$v^\top P_{\mathcal{C}} v = \sum_{t+u \leq \ell} S_t(\mathcal{C}) S_u(\mathcal{C}) v_t v_u \quad \text{and} \quad \|v\|^2 = \sum_t S_t(\mathcal{C}) v_t^2. \tag{21}$$

Let w be the size ℓ vector defined by $w_t = \sqrt{S_t(\mathcal{C})} v_t$. Rewriting the above expression in terms of w yields

$$v^\top P_{\mathcal{C}} v = \sum_{t+u \leq \ell} \sqrt{S_t S_u} w_t w_u \quad \text{and} \quad \|v\|^2 = \|w\|^2, \tag{22}$$

where we omit the dependency of S_t in $\mathcal{C}$.

As a result, the spectral radius of $P_{\mathcal{C}}$ is equal to that of the $\ell \times \ell$ matrix $Q_{\mathcal{C}}$ defined by:

$$Q_{\mathcal{C}} = \begin{pmatrix} S_0 & \sqrt{S_0 S_1} & \cdots & \sqrt{S_0 S_{\ell-1}} & \sqrt{S_0 S_\ell} \\ \sqrt{S_0 S_1} & S_1 & \cdots & \sqrt{S_1 S_{\ell-1}} & 0 \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ \sqrt{S_0 S_{\ell-1}} & \sqrt{S_1 S_{\ell-1}} & \cdots & 0 & 0 \\ \sqrt{S_0 S_\ell} & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

We now finally use the row sum bound to get:

$$\rho(P_{\mathcal{C}}^{(\ell)}) = \rho(Q_{\mathcal{C}}) \leq \max_t \sum_{u \leq \ell - t} \sqrt{S_t S_u} \tag{23}$$

$$\leq \max_t \sum_{u \leq \ell - t} \log(n) \alpha^{\frac{t+u}{2}} \quad \text{via lemma 36} \tag{24}$$

$$= O(\log(n)\alpha^{\ell/2}). \tag{25}$$

Combining the above inequality with Lemma 19 and inequality (20) eventually leads to

$$\rho(\Delta^{(\ell)}) = \widetilde{O}(\alpha^{\ell/2}), \tag{26}$$

which completes the proof of Proposition 17.

3.2. Stability to graph perturbation

In this subsection, we sketch the proofs for Theorems 11 and 12.

A note about computational complexity In the original algorithm, the computation of $B^{(\ell)}$ in polynomial time relies on the almost tree-like, tangle-free structure of the random graph $\mathcal{G}$; this structure may be lost when we add cliques, and increase the algorithm complexity. As we want to devise polynomial algorithms in every case, this may be a hindrance.

Conversely, the computation of the distance matrix $D^{(\ell)}$ can be done in polynomial time (for example breadth-first search of the ℓ-neighbourhood of each vertex in G yields an algorithm in $O(n^{1+\kappa}) = O(n^{13/12})$ in the case of SBM, $O(n^2)$ in general) for any graph, which makes it all the more adapted to the problem at hand.

In order to prove Theorem 11, we need a less restrictive version of Proposition 17; indeed, bounding the spectral radius of the perturbation by $O(\alpha^{\ell/2})$ not only preserves the highest eigenvalues, but also bounds the remaining eigenvalues of $D^{(\ell)}$ by $\sqrt{\lambda_1(D^{(\ell)})}$. This bound is commonly referred to as a Ramanujan-like property of G.

This property, although interesting on its own, is not specifically needed for the reconstruction algorithm to work; rather, we only need one eigenvector associated to the second highest eigenvalue μ_2 to remain unchanged.

We'll therefore only need the following proposition:

Proposition 20 *We consider the same setting as Theorem 11. Let $D^{(\ell)}$ be the distance matrix of G, and $\tilde{G}$ and $\widetilde{D}^{(\ell)}$ be the perturbed versions (after adding adversarial noise) of G and $D^{(\ell)}$, respectively.*

Then

$$\rho(\widetilde{D}^{(\ell)} - D^{(\ell)}) = o(\mu_2^{\ell}). \tag{27}$$

The proof relies on a bound similar to the one in Theorem 10, replacing matrices $P_{\mathcal{C}}$ and $Q_{\mathcal{C}}$ by matrices $P_{\mathcal{K}}$ and $Q_{\mathcal{K}}$ also depending only on the distance to the perturbed vertex set $\mathcal{K}$. The details can be found in the appendix.

References

E. Abbe and C. Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic bp, and the information-computation gap. In *NIPS'16*, 2016.

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *CoRR*, abs/1703.10146, 2017. URL http://arxiv.org/abs/1703.10146.

Emmanuel Abbe, Enric Boix, Peter Ralli, and Colin Sandon. Graph powering and spectral robustness. *CoRR*, abs/1809.04818, 2018. URL http://arxiv.org/abs/1809.04818.

Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: Community detection and non-regular ramanujan graphs. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1347–1357, 2015. doi: 10.1109/FOCS.2015.86. URL https://doi.org/10.1109/FOCS.2015.86.

Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106 (1–19), Dec 2011. doi: 10.1103/PhysRevE.84.066106.

Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75 – 174, 2010. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2009.11.002. URL http://www.sciencedirect.com/science/article/pii/S0370157309002841.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption: clustering sparse networks. *CoRR*, abs/1306.5550, 2013. URL http://arxiv.org/abs/1306.5550.

Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan. Learning communities in the presence of errors. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, pages 1258–1291, 2016. URL http://jmlr.org/proceedings/papers/v49/makarychev16.html.

Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*, STOC '14, pages 694–703, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2710-7. doi: 10.1145/2591796.2591857. URL http://doi.acm.org/10.1145/2591796.2591857.

Ankur Moitra, William Perry, and Alexander S. Wein. How robust are reconstruction thresholds for community detection? In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 828–841, 2016. doi: 10.1145/2897518.2897573. URL https://doi.org/10.1145/2897518.2897573.

A. Montanari and S. Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *STOC'16 Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 2016.

E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *arxiv:1311.4115*, 2013.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2015. ISSN 0178-8051.

Hermann Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.*, 71 (4):441–479, 1912. ISSN 0025-5831. doi: 10.1007/BF01456804. URL https://doi.org/10.1007/BF01456804.

Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. ISSN 0006-3444. doi: 10.1093/biomet/asv008. URL https://doi.org/10.1093/biomet/asv008.

P. Zhang. Robust spectral detection of global structures in the data by learning a regularization. In *NIPS'16*, 2016.

Appendix A. Proof or Propositions 14 and 15

A.1. Outline of the proof and similarities with Bordenave et al. (2015)

The main arguments of the proof rely on the study of three quantities:

(i) a multi-type branching process Z_t,

(ii) a similar process based on exploring the neighbourhood of a vertex v in G, named $Y_t(v)$,

(iii) the actual vectors we're aiming to study, $B^{(\ell)}\chi_k$.

When the ℓ-neighbourhood of v is cycle-free, we have that $B^{(\ell)}\chi_k = \langle \phi_k, Y_t(v)\rangle$ for $k \in [r_0]$; and there is a coupling between the laws of Z_t and $Y_t(v)$ for almost every v, which allows us to translate results on Z_t to results on $B^{(\ell)}\chi_k$.

The proof in Bordenave et al. (2015) studies the matrix B^ℓ, where B is the non-backtracking matrix; B^ℓ_{ij} therefore counts the number of non-backtracking walks between i and j. When the ℓ-neighbourhood of i is tree-like, $(B^{(\ell)}\chi_k)_i = (B^\ell \bar{\chi}_k)i$, where $\bar{\chi}_k$ is a similarly defined vector; most of the results from Bordenave et al. (2015) can therefore be applied to this setting without further work. We will simply lay out the main steps of the proof, highlighting the main differences with Bordenave et al. (2015) when necessary.

A.2. Local structure of G

For an integer $t \geq 0$, we introduce the vector $Y_t(v) = (Y_t(v)(i))_{i\in[r]}$, where

$$Y_t(v)(i) = \left|\{w \in V \mid d(v,w) = t, \sigma(w) = i\}\right|.$$

The proof of our first proposition, although quite lengthy, is completely identical to its equivalent in Bordenave et al. (2015); we therefore omit it.

Proposition 21 *Let $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < 1/8$; then, for all $\gamma < 1/2$:*

(i) for any $k \in [r_0]$, there exists $\rho_k > 0$ such that in probability,

$$\frac{1}{n}\sum_{v\in V} \frac{\langle \phi_k, Y_\ell(v)\rangle^2}{\mu_k^{2\ell}} \to \rho_k.$$

(ii) for any $j \neq k \in [r]$,

$$\mathbb{E}\left|\frac{1}{n}\sum_{v\in V} \langle \phi_j, Y_\ell(v)\rangle\langle \phi_k, Y_\ell(v)\rangle\right| = O\left(\alpha^{5\ell/2} n^{-\gamma/2}(\log(n))^{5/2}\right).$$

(iii) for any $j \neq k \in [r]$,

$$\mathbb{E}\left|\frac{1}{n}\sum_{v\in V} \langle \phi_j, Y_{2\ell}(v)\rangle\langle \phi_k, Y_\ell(v)\rangle\right| = O\left(\alpha^{7\ell/2} n^{-\gamma/2}(\log(n))^{5/2}\right).$$

For $t \geq 0$, define $\mathcal{Y}_t(v) = \{w \in V \mid d(v,w) = t\}$; for $k \in [r]$, we set

$$P_{k,\ell}(v) = \sum_{t=0}^{\ell-1} \sum_{w \in \mathcal{Y}_t(v)} L_k(w),$$

where

$$L_k(w) = \sum_{(x,y)\in \mathcal{Y}_1(w)\setminus \mathcal{Y}_t(v), x\neq y} \langle \phi_k, \tilde{Y}_t(x)\rangle \tilde{S}_{\ell-t-1}(y),$$

$\tilde{Y}_t(x)$ is the equivalent of $Y_t(x)$ when all vertices in $(G,v)_t$ (i.e. vertices at distance at most t from v) are removed and $\tilde{S}_{\ell-t-1}(y) = \|\tilde{Y}_{\ell-t-1}(y)\|_1$.

It can be seen from Bordenave et al. (2015) that when $(G,v)_{2\ell}$ is a tree, then

$$(B^{(\ell)}B^{(\ell)}\chi_k)_v = P_{k,\ell}(v) + \chi_k(v)S_\ell(v) + \langle \phi_k, Y_{2\ell}(v)\rangle.$$

One main difference with the proof in Bordenave et al. (2015) is the presence of the last term in the above sum, as well as the fact that dealing with $B^{(\ell)}B^{(\ell)}\chi_k$ is a little more difficult. The next proposition is an adaptation of Proposition 38 from Bordenave et al. (2015), with an identical – and thus omitted – proof:

Proposition 22 *Let $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < 1/10$. Then, for all $\gamma < 1/2$:*

(i) for all $k \in [r_0]$, there exists ρ'_k such that w.h.p

$$\frac{1}{n}\sum_{v\in V} \frac{(P_{k,\ell}(v) + \langle \phi_k, Y_{2\ell}(v)\rangle)^2}{\mu_k^{4\ell}} \to \rho'_k.$$

(ii) for any $j \neq k \in [r]$, for some $c > 0$:

$$\frac{1}{n}\sum_{v\in V} P_{k,\ell}(v)\langle \phi_j, Y_\ell(v)\rangle = O\left(\alpha^{7\ell/2} n^{-\gamma/2} (\log(n))^c\right).$$

A.3. From local neighbourhoods to the matrix $\mathbf{B}^{(\ell)}$

For ease of notation, we define $N_{k,\ell}(v) = \langle \phi_k, Y_\ell(v)\rangle$; using the same methods as in Bordenave et al. (2015), we have the following estimates:

Proposition 23 *Let $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < 1/4$. Then w.h.p:*

$$\|B^{(\ell)}\chi_k - N_{k,l}\| = o(\alpha^{\ell/2}\sqrt{n}) \text{ and } \|B^{(\ell)}B^{(\ell)}\chi_k - P_{k,\ell} - N_{k,2\ell}\| = O(\alpha^\ell \sqrt{n}).$$

It then remains to follow the proof of Proposition 19 from Bordenave et al. (2015); we simply highlight the proof for estimation (iii) of Proposition 14, since it is the only difference:
Proof (Proposition 14-(iii)): We have by definition

$$\langle \varphi_j, \zeta_k\rangle = \frac{\langle B^{(\ell)}\chi_j, B^{(\ell)}B^{(\ell)}\chi_k\rangle}{\|B^{(\ell)}\chi_j\|\|B^{(\ell)}B^{(\ell)}\chi_k\|}.$$

But $\|B^{(\ell)}\chi_j\| = \Theta(\sqrt{n}\mu_k^\ell)$, $\|B^{(\ell)}B^{(\ell)}\chi_k\| = \|B^{(\ell)}\chi_k\|\theta_k = \Theta(\sqrt{n}\mu_k^{2\ell})$ and:

$$\begin{aligned}\left|\langle B^{(\ell)}\chi_j, B^{(\ell)}B^{(\ell)}\chi_k\rangle - \langle N_{j,\ell}, P_{k,\ell} + N_{k,2\ell}\rangle\right| &\leq \|N_{j,\ell}\|\|B^{(\ell)}B^{(\ell)}\chi_k - P_{k,\ell} - N_{k,2\ell}\| \\ &\quad + \|B^{(\ell)}B^{(\ell)}\chi_k\|\|B^{(\ell)}\chi_j - N_{j,\ell}\| \\ &= \widetilde{O}(\alpha^{4\ell}\sqrt{n}).\end{aligned}$$

Furthermore, from Propositions 21 and 22, we get

$$\langle N_{j,\ell}, P_{k,\ell} + N_{k,2\ell}\rangle = \widetilde{O}(\alpha^{7\ell/2}n^{1-\gamma/2}).$$

This gives the desired result. ■

A.4. Ramanujan property of $B^{(\ell)}$

In order to complete the proof of Theorem 7, we need a control on the other eigenvalues of $B^{(\ell)}$. This is covered by the following proposition:

Proposition 24 *Let $H = \langle\varphi_1, \dots, \varphi_{r_0}\rangle$, and $\ell \sim \kappa\log_\alpha(n)$ with $\kappa < 1/12$. Then with high probability*

$$\sup_{x\in H^\perp, \|x\|=1} \|B^{(\ell)}x\| = \widetilde{O}(\alpha^{\ell/2}). \tag{28}$$

The proof of this result relies on the following decomposition of $B^{(\ell)}$, whose proof can be found in Massoulié (2014):

Lemma 25 *Matrix $B^{(\ell)}$ verifies the identity*

$$B^{(\ell)} = \Delta^{(\ell)} + \sum_{m=1}^{\ell} \Delta^{(\ell-m)}\bar{A}B^{(m-1)} - \sum_{m=0}^{\ell}\Gamma^{\ell,m}, \tag{29}$$

for matrices $\Delta^{(j)}, \Gamma^{\ell,m}$ such that for $\ell = O(\log(n))$ and with high probability, for all $\varepsilon > 0$,

$$\rho(\Delta^{(j)}) = \widetilde{O}(\alpha^j),\ j = 1, \dots, \ell, \tag{30}$$

$$\rho(\Gamma^{\ell,m}) = n^{\varepsilon-1}\alpha^{(\ell+m)/2}, m = 1, \dots, \ell. \tag{31}$$

Here, $\bar{A}$ refers to the expected value of the adjacency matrix A of G.

The next step is therefore to control $B^{(m-1)}x$ for $x \in H^\perp$; in what follows γ will be any constant below $1/2$. We begin with the following proposition from Bordenave et al. (2015):

Proposition 26 *Let $\ell \sim \kappa\log_\alpha(n)$ with $\kappa < \gamma/2$. There exists a subset $\mathcal{B} \subset V$, constants C and c such that w.h.p the following holds:*

(i) for all $i \in V \setminus \mathcal{B}$, $0 \leq m \leq \ell$,

$$|(B^{(m)}\chi_k)_i - \mu_k^{t-\ell}(B^{(\ell)}\chi_k)_i| \leq C\log(n)^c\alpha^{m/2} \qquad \text{if } k \in [r_0],$$

$$|(B^{(m)}\chi_k)_i| \leq C\log(n)^c\alpha^{m/2} \qquad \text{if } k \in [r] \setminus [r_0].$$

(ii) for all $i \in \mathcal{B}$, $0 \leq m \leq \ell$ and $k \in [r]$,

$$|(B^{(\ell)}\chi_k)_i| \leq C\log(n)^c\alpha^m.$$

(iii) $|\mathcal{B}| = \widetilde{O}(\alpha^\ell n^{1-\gamma})$.

From this, we get the following corollary:

Corollary 27 *Let $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < \gamma/2$; then, with high probability, for $0 \leq m \leq \ell - 1$ and $k \in [r_0]$:*

$$\sup_{x \perp B^{(\ell)}\chi_k, \|x\|=1} \langle B^{(m)}\chi_k, x\rangle = \widetilde{O}(\sqrt{n}\,\alpha^{m/2}).$$

Additionally, for $k \in [r] \setminus [r_0]$,

$$\|B^{(m)}\chi_k\| = \widetilde{O}(\sqrt{n}\alpha^{m/2}).$$

Proof Write

$$\langle B^{(m)}\chi_k, x\rangle = \sum_{i\in\mathcal{B}} x_i(B^{(m)}\chi_k)_i + \sum_{i\notin\mathcal{B}} x_i(B^{(m)}\chi_k)_i = s_1 + s_2.$$

Using the Cauchy-Schwarz inequality, the first sum is bounded by

$$|s_1| \leq \log(n)^c\alpha^m\sqrt{|\mathcal{B}|} \leq \log(n)^d\alpha^m\alpha^{\ell/2}n^{(1-\gamma)/2} = o(\sqrt{n}\alpha^{m/2}),$$

while the second can be treated using Proposition 26 and the fact that $\langle B^{(\ell)}\chi_k, x\rangle = 0$:

$$\begin{aligned}
|s_2| &\leq \mu_k^{t-\ell}\sum_{i\in\mathcal{B}}|x_i||(B^{(\ell)}\chi_k)_i| + \sum_{i\notin\mathcal{B}}|x_i||(B^{(m)}\chi_k)_i - \mu_k^{t-\ell}(B^{(\ell)}\chi_k)_i| \\
&\leq \log(n)^c\alpha^{t-\ell}\alpha^\ell\alpha^{\ell/2}n^{(1-\gamma)/2} + \log(n)^c\sqrt{n}\alpha^{t/2} \\
&= \widetilde{O}(\sqrt{n}\,\alpha^{m/2}),
\end{aligned}$$

where we used again the Cauchy-Schwarz inequality as before.

Let now $k \in [r] \setminus [r_0]$; as before, we write

$$\begin{aligned}
\|B^{(m)}\chi_k\|^2 &= \sum_{i\in\mathcal{B}}(B^{(m)}\chi_k)_i^2 + \sum_{i\notin\mathcal{B}}(B^{(m)}\chi_k)_i^2 \\
&\leq |\mathcal{B}|\log(n)^c\alpha^{2m} + n\log(n)^c\alpha^m \\
&= n\log(n)^c(\alpha^{l+2m}n^{-\gamma} + \alpha^m) \\
&= \widetilde{O}(n\alpha^m),
\end{aligned}$$

and the result follows. ■

We are now ready to prove Proposition 24:

Proof Let $x \in H^\perp$ such that $\|x\| = 1$ and the supremum in (28) is reached; using the decomposition from Lemma 25, we have

$$\|B^{(\ell)}x\| \le \rho(\Delta^{(\ell)}) + \sum_{m=1}^{\ell} \rho(\Delta^{(\ell-m)})\|\bar{A}B^{(m-1)}x\| + \sum_{m=1}^{\ell} \rho(\Gamma^{\ell,m}).$$

The first and third terms are bounded by $\widetilde{O}(\alpha^{\ell/2})$. For the second term, we notice that defining the matrix P by

$$P = \frac{1}{n}\sum_{k=1}^{r} \mu_k \chi_k \chi_k^\top,$$

we have $\bar{A} = P - \mathrm{diag}(P)$ since $W = \sum \mu_k \phi_k \phi_k^\top$.

Therefore, for fixed $1 \le m \le \ell$, we have:

$$\begin{aligned}\|\bar{A}B^{(m-1)}x\| &= \left\| \sum_{k=1}^{r} \mu_k \chi_k \chi_k^\top B^{(m-1)}x - \mathrm{diag}(P)B^{(m-1)}x \right\| \\ &\le \frac{\sup_i W_{ii}}{n}\|B^{(m-1)}x\| + \sum_{k\in[r_0]} \frac{\mu_k}{n}\|\chi_k\chi_k^\top B^{(m-1)}x\| + \sum_{k\in[r]\setminus[r_0]} \frac{\mu_k}{n}\|\chi_k\chi_k^\top B^{(m-1)}x\| \\ &= I + J + K.\end{aligned}$$

Notice first that $B^{(\ell)}_{ij} \le 2$ for all i, j by the tangle-free property, so $I = O(1)$. Now, for $k \in [r_0]$, we have

$$\begin{aligned}\|\chi_k\chi_k^\top B^{(m-1)}x\| &= \|\chi_k\|\langle B^{(m-1)}\chi_k, x\rangle \\ &\le \widetilde{O}(\sqrt{n} \times \sqrt{n}\alpha^{m/2}).\end{aligned}$$

Therefore, $J = \widetilde{O}(\alpha^{m/2})$; finally, using the Cauchy-Schwarz inequality, we have for $k \in [r] \setminus [r_0]$

$$\begin{aligned}\|\chi_k\chi_k^\top B^{(m-1)}x\| &\le \|\chi_k\|\|B^{(m-1)}\chi_k\|\|x\| \\ &= \widetilde{O}(\sqrt{n} \times \sqrt{n}\alpha^{m/2} \times 1).\end{aligned}$$

Putting this all together, we find that for $1 \le m \le \ell$

$$\|\bar{A}B^{(m-1)}x\| = \widetilde{O}(\alpha^{m/2}).$$

Since $\rho(\Delta^{(\ell-m)}) = \widetilde{O}(\alpha^{(\ell-m)/2})$, we get $\|B^{(\ell)}x\| = \widetilde{O}(\alpha^{\ell/2})$, which proves the desired result.
■

A.5. Proof of Proposition 15

Using Proposition 24, we are now able to prove our last result. Note that if $\kappa < 1/12$, there exists a $\gamma < 1/2$ such that $\kappa < \gamma/6$.

Let z_k be the Gram-Schmidt orthonormalization of φ_k; using Lemma 9 from Bordenave et al. (2015), we know that

$$\|\varphi_k - z_k\| = \widetilde{O}(\alpha^{3\ell/2} n^{-\gamma/2}),$$

and thus z_k is asymptotically parallel to φ_k.

We only need a final lemma to complete our proof:

Lemma 28 *Assume that $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < \gamma/6$. Then*

$$\|\zeta_k - z_k\| = \widetilde{O}(\theta_k^{-1} \alpha^{\ell/2}).$$

Proof Write

$$\zeta_k = \sum_{j \in [r_0]} \langle \zeta_k, z_j \rangle z_j + x,$$

where $x \in H^\perp$.

We have, for $j \neq k$, $\langle \zeta_k, z_j \rangle = \widetilde{O}(\alpha^{2\ell} n^{-\gamma/2})$ by the above bound of $\|\varphi_j - z_j\|$; furthermore,

$$\begin{aligned} \|x\|^2 &= \langle \zeta_k, x \rangle \\ &= \theta_k^{-1} \langle B^{(\ell)} \varphi_k, x \rangle \\ &\leq \theta_k^{-1} \|B^{(\ell)} x\| \\ &= \widetilde{O}(\theta_k^{-1} \alpha^{\ell/2}) \times \|x\|. \end{aligned}$$

Therefore, we can write

$$\begin{aligned} 1 = \|\zeta_k\|^2 &= \langle \zeta_k, z_k \rangle^2 + \sum_{j \neq k} \langle \zeta_k, z_j \rangle^2 + \|x\|^2 \\ &= \langle \zeta_k, z_k \rangle^2 + \widetilde{O}(\alpha^{2\ell} n^{-\gamma/2}) + \widetilde{O}(\theta_k^{-2} \alpha^{\ell}) \\ &= \langle \zeta_k, z_k \rangle^2 + \widetilde{O}(\theta_k^{-2} \alpha^{\ell}), \end{aligned}$$

since $\kappa < \gamma/6$.

Then,

$$\|z_k - \zeta_k\|^2 = 2(1 - \langle \zeta_k, z_k \rangle) = \widetilde{O}(\theta_k^{-2} \alpha^{\ell}),$$

which yields the desired result. ■

Proof (of Proposition 15): We first bound $\|B^{(\ell)} z_k - D z_k\|$ for $k \in [r_0]$. Notice that $D z_k = \theta_k z_k$; this gives

$$\begin{aligned} \|B^{(\ell)} z_k - D z_k\| &\leq \|B^{(\ell)} z_k - B^{(\ell)} \varphi_k\| + \|B^{(\ell)} \varphi_k - \theta_k z_k\| \\ &\leq \rho(B^{(\ell)}) \|z_k - \varphi_k\| + \theta_k \|\zeta_k - z_k\| \\ &= O(\alpha^{\ell}) \times \widetilde{O}(\alpha^{3\ell/2} n^{-\gamma/2}) + \widetilde{O}(\alpha^{\ell/2}) \\ &= \widetilde{O}(\alpha^{\ell/2}). \end{aligned}$$

Consider now $x \in \mathbb{R}^V$ such that $\|x\| = 1$. Decomposing x as $\sum x_k z_k + x'$ where $x' \in H^\perp$, we have:

$$\begin{aligned}\|B^{(\ell)}x - Dx\| &\leq \sum_{k\in[r_0]} x_k \|B^{(\ell)}z_k - Dz_k\| + \|B^{(\ell)}x' - Dx'\| \\ &\leq \widetilde{O}(\alpha^{\ell/2}) + \|B^{(\ell)}x'\| \\ &= \widetilde{O}(\alpha^{\ell/2}),\end{aligned}$$

which completes the proof. ■

Appendix B. Proofs for Theorem 9

B.1. Proof of Lemma 16

We first recall a result from Kesten and Stigum: consider a multitype Galton-Watson process, where the type of the root node is distributed according to arbitrary probability vector ν, and a particle of type $j \in [r]$ has a $\mathrm{Poi}(M_{ij})$ number of children of type i. Let Z_t be the vector of population at time t, and $\mathcal{F}_t$ the natural filtration associated to Z_t; we have the following statement:

Lemma 29 *For each μ eigenvalue of M such that $\mu^2 > \alpha$, and each eigenvector ϕ associated to μ,*

$$t \mapsto X(\phi, \nu, t) = \mu_k^{-t}\langle \phi_k, Z_t\rangle \tag{32}$$

is an $\mathcal{F}_t$-martingale converging a.s. and in L^2 to a random variable with finite variance and expected value $\langle \phi_k, \nu\rangle$.

Let $\mu \neq \alpha$ be an eigenvalue of M of multiplicity d such that $\mu^2 > \alpha$, and $\phi^{(1)}, \dots, \phi^{(d)}$ an orthonormal basis of eigenvectors associated to d. We define for all $i \in [d], j \in [r]$, $X_j^{(i)}$ the limit variable of martingale (32), applied to $\phi = \phi^{(i)}$ and $\nu = \delta_j$. Similarly to previous notations, let $\phi^{(i)}$ (resp. $X^{(i)}$) be the vector $\left(\phi_j^{(i)}\right)_{j\in[r]}$ (resp. $\left(X_j^{(i)}\right)_{j\in[r]}$), and $\boldsymbol{\phi}$ (resp. $\mathbf{X}$) the (random) matrix whose columns are the $\phi^{(i)}$ (resp. the $X^{(i)}$). Recall that from Lemma 29, the expected value of $X_j^{(i)}$ is $\phi_j^{(i)}$ for all i, j.

Now, let ξ be an eigenvector of $B^{(\ell)}$, normalized so that $\|\xi\|^2 = n$, with associated eigenvalue $\Theta(\mu^\ell)$; as shown in the proof of Theorem 7, there exists a vector $u \in \mathbb{R}^d$ such that

$$\|\xi - (\langle \boldsymbol{\phi} u, Y_\ell(v)\rangle)_{v\in V}\| = o(1).$$

We let

$$\phi^{(\xi)} = \boldsymbol{\phi} u \quad \text{and} \quad X^{(\xi)} = \mathbf{X} u.$$

From Proposition 21, u has norm $\Theta(1)$, and since $\mu^{-t}\langle \phi, Z_t\rangle$ (with $\nu = \delta_j$) converges to $X_j^{(i)}$ in L^1, $\mu^{-t}\langle \phi^{(\xi)}, Z_t\rangle$ converges to $X_j^{(\xi)}$ in L^1 independently of ξ.

Using proposition 36 from Bordenave et al. (2015), we have the following:

Lemma 30 *For all $i \in [r]$, we have the following convergence in L^1:*

$$\frac{1}{n}\sum_{v\in V} \mathbf{1}_{\sigma(v)=i}\,\xi(v)\mathbf{1}_{|\xi(v)|\le K} \to \pi(i)\mathbb{E}\Big[X^{(\xi)}\mathbf{1}_{|X^{(\xi)}|\le K}\Big],$$

for all K that is a continuity point of the distribution of X_i, and independently of ξ.

Proof We first recall the aforementioned proposition from Bordenave et al. (2015): we say that a function τ that takes a graph and a distinguished vertex as an argument is ℓ-local if $\tau(G, v)$ depends only on the ℓ-neighbourhood of v in G. Denote by T the multitype Galton-Watson tree discussed earlier, rooted at o, where o has the distribution δ_ι and ι has distribution π.

Lemma 31 *Assume that τ, ψ are two ℓ-local functions such that $|\tau| \le \psi$ and ψ is non-decreasing by the addition of edges. Then, if $\ell \sim \kappa \log_\alpha(n)$ with $\kappa < 1/2$, we have, for $\gamma < 1/2$:*

$$\begin{aligned}\mathbb{E}&\left[\left|\frac{1}{n}\sum_{v\in V}\tau(G,v) - \mathbb{E}[\tau(T,o)]\right|\right] \\ &\le c\frac{\alpha^{\ell/2}\sqrt{\log(n)}}{n^{\gamma/2}}\left(\left(\mathbb{E}[\max_{v\in V}\psi^4(G,v)]\right)^{1/4} \vee \left(\mathbb{E}\big[\psi^2(T,o)\big]\right)^{1/2}\right)\end{aligned}$$

We now apply this lemma with $\tau(G, v) = \mathbf{1}_{\sigma(v)=i}\,\langle\phi^{(\xi)}, Y_\ell(v)\rangle\mathbf{1}_{|\langle\phi^{(\xi)},Y_\ell(v)\rangle|\le K}$ where Y_ℓ is defined in Proposition 21.
We can set $\psi(G, v) = K$, and by Lemma 29 and the subsequent analysis, we have

$$\mathbb{E}[\tau(T,o)] \to \pi(i)\,\mathbb{E}\Big[X_i^{(\xi)}\mathbf{1}_{|X_i^{(\xi)}|\le K}\Big]$$

independently of ξ.

Now, by definition, we have $\|\xi - \big(\langle\phi^{(\xi)}, Y_\ell(v)\rangle\big)_{v\in V}\| = o(1)$ (again independently of ξ). By the Cauchy-Schwarz inequality, we deduce that

$$\frac{1}{n}\sum_{v\in V}\Big|\xi(v) - \langle\phi^{(\xi)}, Y_\ell(v)\rangle\Big| = o(1)$$

as well, and the lemma follows if K is a continuity point of $X_i^{(\xi)}$. ∎

It now remains to prove the desired properties of the $X_i^{(\xi)}$; first, since $\mu \neq \alpha$, then $\phi^{(\xi)}$ is orthogonal to the all-one vector and as such

$$\sum_{i\in[r]}\mathbb{E}[X_i] = \sum_{i\in[r]}\phi_i^{(\xi)} = 0$$

Moreover, $\|\phi^{(\xi)}\|^2 = \|u\|^2 = \Theta(1)$, which proves the second assertion.

Finally, let $\eta > 0$; since the $X_j^{(i)}$ all have finite variance, there exists a constant $K > 0$ such that $\mathbb{P}(|X_j^{(i)}| \le K') \ge 1 - \eta$ for all i, j and thus

$$\mathbb{P}\big(\|\mathbf{X}\|_\infty \le K'\big) \ge 1 - dr\eta.$$

Using the equivalence of norms, we find

$$\mathbb{P}\Big(\forall i,\ \|X^{(i)}\|_2^2 \leq rK'^2\Big) \geq 1 - dr\eta$$

which implies (since $X^{(\xi)} = \mathbf{X}u$)

$$\mathbb{P}\Big(\|X^{(\xi)}\|_2^2 \leq r\|u\|_2^2 K'^2\Big) \geq 1 - dr\eta.$$

Using again norm equivalence yields finally, for $K = \sqrt{r}\|u\|K'$,

$$\mathbb{P}\Big(\|X^{(\xi)}\|_\infty \leq K\Big) \geq 1 - dr\eta. \tag{33}$$

Now, we have, for all i,

$$\begin{aligned}\Big|\mathbb{E}[X_i^{(\xi)}\mathbf{1}_{|X_i^{(\xi)}|\leq K}] - \mathbb{E}[X_i^{(\xi)}]\Big| &= \mathbb{E}\Big[X^{(\xi)_i}\mathbf{1}_{|X_i^{(\xi)}|>K}\Big] \\ &\leq \sqrt{\mathbb{E}\Big[(X_i^{(\xi)})^2\Big]\mathbb{P}\Big(|X_i^{(\xi)}| \geq K\Big)} \\ &\leq \sqrt{\mathbb{E}\Big[(X_i^{(\xi)})^2\Big]} \cdot \sqrt{dr\eta}\end{aligned}$$

But by Doob's Theorem, $\mathbb{E}\Big[(X_i^{(\xi)})^2\Big]$ is finite so choosing η accordingly yields the last inequality of Lemma 16.

B.2. Proof of limit (15)

For each $v \in V$, we define I_v to be the random variable equal to 1 if v is assigned to I^+, and 0 otherwise. Conditionnally to ξ, it is straightforward to see that

$$I_v \sim \mathrm{Ber}(q_v) \quad \text{with} \quad q_v = \frac{1}{2} + \frac{1}{2K}\xi(v)\mathbf{1}_{|\xi(v)|\leq K}$$

Now, let $P_i = (|\{v \in I^+ \mid \sigma(v) = i\}|)/n$; by definition,

$$P_i = \frac{1}{n}\sum_{v\in V}\mathbf{1}_{\sigma(v)=i}\, I_v$$

.

We therefore have

$$\mathbb{E}[P_i \mid \xi] = \frac{1}{n}\sum_{v\in V}\mathbf{1}_{\sigma(v)=i}\left(\frac{1}{2} + \frac{1}{2K}\xi(v)\mathbf{1}_{|\xi(v)|\leq M}\right), \quad \mathrm{Var}(P_i) \leq \frac{1}{4n}$$

and thus with high probability, independently of ξ,

$$\begin{aligned}P_i &= \mathbb{E}[P_i \mid \xi] + n^{-1/3} &(34)\\ &\to \pi(i)\left(\frac{1}{2} + \frac{1}{2K}\,\mathbb{E}[X_i^{(\xi)}\mathbf{1}_{|X_i^{(\xi)}|\leq M}]\right) = \pi(i)\,\tilde{p}_i &(35)\end{aligned}$$

where the convergence speed is independent from ξ.

B.3. Explicit bounds on K

In this section, the goal is to perform a more precise analysis of the limit variables $X_i^{(\xi)}$, and to leverage this analysis to obtain an explicit value for K in Algorithm 8. For simplicity, we will assume that $\pi \equiv 1/r$ throughout this section, although most of the results hold for general π. We begin with a small lemma:

Lemma 32 *Let ϕ be a normed eigenvector of M associated to an eigenvalue $\alpha > \mu > \sqrt{\alpha}$, and denote by $X_i^{(\phi)}$ the limit random variables of Lemma 29. Let $c^{(\phi),j}$ (resp. $m^{(\phi),j}$) the vector of the j-th cumulants (resp. moments) of the $X_i^{(\phi)}$. We then have the following relation, for all $j \in \mathbb{N}$:*

$$c^{(\phi),j} = \frac{M}{\mu^j} m^{(\phi),j}$$

By definition, we have $c^{(\phi),1} = m^{(\phi),1} = \phi$, and we have the following corollary for $c^{(\phi),2}$ and $m^{(\phi),2}$:

Corollary 33 *We denote by ϕ^2 the vector whose coordinates are the ϕ_i^2. Then*

$$c^{(\phi),2} = \left(I - \frac{M}{\mu^2}\right)^{-1} \frac{M}{\mu^2}\,\phi^2 \quad \textit{and} \quad m^{(\phi),2} = \left(I - \frac{M}{\mu^2}\right)^{-1} \phi^2.$$

As a result, we have

$$\sum_{i\in[r]} \mathrm{Var}(X_i^{(\phi)}) = \frac{1}{\tau - 1} \quad \textit{and} \quad \sum_{i\in[r]} \mathbb{E}[(X_i^{(\phi)})^2] = \frac{\tau}{\tau - 1},$$

where $\tau = \mu_2^2/\alpha$.

Proof (of Corollary 33). The first part is an easy calculation, observing that $c^{(\phi),2} = m^{(\phi),2} - \phi^2$.

For the second part, since the all-one vector e is an eigenvector of M associated to the eigenvalue α, we have:

$$\begin{aligned}
\sum_{i\in[r]} \mathrm{Var}(X_i^{(\phi)}) &= \langle e, c^{(\phi),2}\rangle \\
&= e^\top \left(I - \frac{M}{\mu^2}\right)^{-1} \frac{M}{\mu^2}\,\phi^2 \\
&= \frac{\alpha/\mu^2}{1-\alpha/\mu^2} e^\top \phi^2 \\
&= \frac{1}{\tau - 1},
\end{aligned}$$

and a similar calculation yields the second identity. ■

It now remains to prove Lemma 32:

Proof (of Lemma 32). Using the Galton-Watson tree definition (and going one step down into the tree), we have the following characterization for the variables $X_i^{(\phi)}$:

$$X_i^{(\phi)} = \frac{1}{\mu} \sum_{j \in [r]} \sum_{k=1}^{\mathrm{Poi(M_{ij})}} X_{j,k}^{(\phi)},$$

where the $X_{j,k}^{(\phi)}$ are independent copies of $X_j^{(\phi)}$ for all k. Applying the Laplace transform (denoted by $\psi_i^{(\phi)}$) and taking the logarithm on both sides, we find that for all $t \in \mathbb{R}$,

$$\log(\psi_i^{(\phi)}(t)) = \sum_{j \in [r]} M_{ij} \left(\psi_j^{(\phi)} \left(\frac{t}{\mu} \right) - 1 \right)$$

Now, the k-th Taylor coefficient of the LHS is $c_i^{(\phi),k}/k!$, and the one on the RHS is

$$\sum_{j \in [r]} M_{ij} \frac{m_j^{(\phi),k}}{k!\,\mu^k} = \frac{1}{\mu^k\,k!} \left[M m^{(\phi),k} \right]_i,$$

which completes the proof. ■

We now can prove our first result on the vector u defined before:

Lemma 34 *Let μ, ξ and u be defined as in the proof of Lemma 16. Then we have*

$$\|u\|^2 = r(\tau - 1) + o(1)$$

Proof From Lemma 21, we know that for each $i \in [d]$,

$$\left\| \left(\langle \phi^{(i)}, Y_\ell(v) \rangle \right)_{v \in V} \right\|^2 = n(\rho^{(i)} + o(1)) \quad \text{where} \quad \rho^{(i)} = \sum_{i \in [r]} \pi(i) \mathbb{E}\left[(X^{(i)})^2 \right]$$

But since $\pi \equiv 1/r$, we know from Corollary 33 that

$$\rho^{(i)} = \rho := \frac{1}{r(\tau - 1)}$$

But the vectors $\left(\langle \phi^{(i)}, Y_\ell(v) \rangle \right)_{v \in V}$ are asymptotically orthogonal, and thus

$$n = \|\xi\|^2 = (\|v\|^2 + o(1)) \cdot n \cdot (\rho + o(1)),$$

which yields the desired result. ■

Now, we are ready to prove some bounds for K; the main step is the following Markov bound on $(X^{(i)})^2$:

Lemma 35 *Let $\eta > 0$; then, for all $i \in [d]$, $j \in [r]$,*

$$\mathbb{P}\left(|X_j^{(i)}| \leq \sqrt{\frac{\tau}{\eta(\tau - 1)}} \right) \geq 1 - \eta$$

Proof For all $C > 0$, we have by Markov's inequality

$$\mathbb{P}\Big(|X_j^{(i)}| \geq C\Big) \leq \frac{\mathbb{E}[(X_j^{(i)})^2]}{C^2} \leq \frac{\tau}{C^2(\tau-1)},$$

where we bounded $\mathbb{E}[(X_j^{(i)})^2]$ by the sum of all $\mathbb{E}[(X_k^{(i)})^2]$. The lemma then follows easily. ■

Now, we have to unravel the calculations done in the proof for Lemma 16; let $\varepsilon < 0$. By the same bound as above (as well as the fact that the $\phi^{(i)}$ are orthogonal), we have

$$\mathbb{E}\Big[(X_i^{(\xi)})^2\Big] \leq \|u\|^2 \frac{\tau}{\tau-1} = r\tau + o(1)$$

Therefore, an asymptotically good choice of η is

$$\eta = \frac{\varepsilon^2}{r^2\, d\, \tau},$$

which yields a value for K' of

$$K' = \sqrt{\frac{\tau}{\eta(\tau-1)}} = \frac{r\,\tau}{\varepsilon}\sqrt{\frac{d}{\tau-1}}$$

Finally, the bound for K becomes

$$K = \sqrt{r}\|u\|K' = \frac{r\,\tau}{\varepsilon}\sqrt{r\,d\,\tau}$$

We thus need to find a sufficient value for ϵ; recall that

$$\sum_{i\in[r]} \mathbb{E}\Big[X_i^{(\xi)}\Big]^2 = \|u\|^2 = r(\tau-1) + o(1) \quad \text{and} \quad \sum_{i\in[r]} \mathbb{E}\Big[X_i^{(\xi)}\Big] = 0,$$

and therefore there exists some values for i, j such that

$$\left|\mathbb{E}\Big[X_i^{(\xi)}\Big]^2 - \mathbb{E}\Big[X_j^{(\xi)}\Big]^2\right| \geq \sqrt{r(\tau-1)} + o(1).$$

A sufficient choice of ε is thus $\sqrt{r(\tau-1)}/4$, which yields an explicit value for K:

$$K = \frac{r\,\tau}{\varepsilon}\sqrt{r\,d\,\tau} = r\,\tau\sqrt{d\,\frac{\tau}{\tau-1}}$$

Appendix C. Proof of Lemma 18

We first recall some results about the neighbourhoods of vertices, whose proofs can be found in Bordenave et al. (2015):

Lemma 36 *For a vertex i, define $S_t(i)$ as the number of vertices at distance t of i.*

Then there exist constants C and $\varepsilon > 0$ such that with probability $1 - O(n^{-\varepsilon})$, for all $i \in \{1, \dots n\}$ and $\ell = O(\log(n))$:

$$S_t(i) \leq C \cdot \log(n) \cdot \alpha^t, \quad t \in \{1, \dots, \ell\}. \tag{36}$$

On the other hand, with high probability, when $\ell = \kappa \log_\alpha(n)$ with $\kappa < 1/2$:

$$\sum_{i=1}^{n} S_\ell(i)^2 = \Theta(n\alpha^{2\ell}). \tag{37}$$

Additionnally, a result about the almost tree-like structure of vertex neighbourhoods:

Lemma 37 *Assume $\ell = \kappa \log(n)$, with $\kappa \log(\alpha) < 1/4$. Then with high probability no node i has more than one edge cycle in its ℓ-neighbourhood; we say that G is ℓ-tangle-free.*

Using those results, we are now able to prove Lemma 18:

Proof From Lemma 37, we can deduce that if $d(i,j) \leq \ell$, there are at most two distinct paths between i and j. Therefore, $B_{ij}^{(\ell)} \leq 2$ for all i, j.

Additionally, if $D_{ij}^{(\ell)} = 1$, then there is a self-avoiding path of length ℓ between i and j, and thus $B_{ij}^{(\ell)} = 1$, so $\Delta_{ij}^{(\ell)} \geq 0$ for all i, j.

Finally, assume that there exists a pair i, j such that $D_{ij}^{(\ell)} = 0$ and $B_{ij}^{(\ell)} = 2$; then there are two paths of length ℓ between i and j and $d(i,j) < \ell$ so there is also a path of length less than ℓ. This contradicts Lemma 37.

Consider now two vertices i and j such that $\Delta_{ij}^{(\ell)} = 1$, there are two possibilities:

(i) $D_{ij}^{(\ell)} = 0$ and $B_{ij}^{(\ell)} > 0$: then $d(i,j) < \ell$ and there is a path of length $< \ell$ and at least a path of length ℓ between i and j.

(ii) $D_{ij}^{(\ell)} = 1$ and $B_{ij}^{(\ell)} > 1$: then there are at least two paths of length ℓ between i and j.

In both cases, there are at least two paths of length at most ℓ connecting i and j, which implies the statement of the lemma. ■

Appendix D. Proof of Lemma 19

Proof (i) is obvious since $d(i, \mathcal{C}) + d(j, \mathcal{C}) \leq \ell$ implies $d(i, \mathcal{C}) \leq \ell$.

For (ii), note first that $V_\mathcal{C}$ and $V_{\mathcal{C}'}$ are disjoint for $\mathcal{C} \neq \mathcal{C}'$: if $i \in V_\mathcal{C} \cap V_{\mathcal{C}'}$, then $\mathcal{C}$ and $\mathcal{C}'$ are in the ℓ-neighbourhood of i, which contradicts Lemma 37.

Let $\pi_\mathcal{C}$ be the projection on $V_\mathcal{C}$ for all $\mathcal{C}$; the $\pi_\mathcal{C}$ are mutually orthogonal and for a vector v, we have:

$$v^\top P^{(\ell)} v = \sum_{\mathcal{C}} v^\top \pi_\mathcal{C} P^{(\ell)} \pi_\mathcal{C} v = \sum_{\mathcal{C}} (\pi_\mathcal{C} v)^\top P_\mathcal{C}^{(\ell)} (\pi_\mathcal{C} v) \tag{38}$$

$$\leq \sum_{\mathcal{C}} \rho(P_\mathcal{C}^{(\ell)}) \cdot \|\pi_\mathcal{C} v\|^2 \tag{39}$$

$$\leq \max_{\mathcal{C}} \rho(P_\mathcal{C}^{(\ell)}) \cdot \sum_{\mathcal{C}} \|\pi_\mathcal{C} v\|^2. \tag{40}$$

On the other hand,

$$\|v\|^2 \geq \sum_{\mathcal{C}} \|\pi_{\mathcal{C}} v\|^2. \tag{41}$$

Combining inequalities (40) and (41) yields $\rho(P^{(\ell)}) \leq \max_{\mathcal{C}} \rho(P_{\mathcal{C}}^{(\ell)})$; the reverse inequality comes from the decomposition $P^{(\ell)} = \sum P_{\mathcal{C}}^{(\ell)}$.

■

Appendix E. Proof of Proposition 20

In the same vein as Lemma 36, for a vertex set $\mathcal{X}$, define $S_t(\mathcal{X})$ as the number of vertices at distance t of $\mathcal{X}$. By taking the union on all vertices of $\mathcal{X}$, we easily get the following corollary:

Corollary 38 *For the same constants C and ε as above, with probability $1 - O(n^{-\varepsilon})$, we have for all vertex subsets $\mathcal{X} \in \mathcal{P}(\{1, \dots, n\})$ and $\ell = O(\log(n))$:*

$$S_t(\mathcal{X}) \leq C \cdot |\mathcal{X}| \log(n) \cdot \alpha^t, \quad t \in \{1, \dots, \ell\}.$$

We are now able to prove Proposition 20:

Proof Let $\mathcal{K}$ be the modified vertex set, and consider vertices i and j such that $D_{ij}^{(\ell)} \neq \widetilde{D}_{ij}^{(\ell)}$. Then we have one of four possibilities:

(i) $\tilde{d}(i,j) = \ell$ and $d(i,j) < \ell$

(ii) $\tilde{d}(i,j) > \ell$ and $d(i,j) = \ell$

(iii) $\tilde{d}(i,j) = \ell$ and $d(i,j) > \ell$

(iv) $\tilde{d}(i,j) < \ell$ and $d(i,j) = \ell$

In cases (i) and (ii), there is a path between i and j in G through $\mathcal{K}$ of length at most ℓ, and in cases (iii) and (iv) there is a path between i and j in $\tilde{G}$ through $\mathcal{K}$. Therefore, in all cases, we have that

$$d(i, \mathcal{K}) + d(j, \mathcal{K}) \leq \ell.$$

Write $|\widetilde{D}^{(\ell)} - D^{(\ell)}|$ for the matrix whose (i,j) coefficient is $|\widetilde{D}_{ij}^{(\ell)} - D_{ij}^{(\ell)}|$, and $P_{\mathcal{K}}$ for the matrix such that $P_{\mathcal{K},ij} = \mathbf{1}\{d(i,\mathcal{K}) + d(j,\mathcal{K}) \leq \ell\}$; the previous analysis and the Perron-Frobenius theorem imply that

$$\rho(\widetilde{D}^{(\ell)} - D^{(\ell)}) \leq \rho(|\widetilde{D}^{(\ell)} - D^{(\ell)}|) \leq \rho(P_{\mathcal{K}}). \tag{42}$$

We can then perform the same analysis as in the proof of Proposition 17 to find that the spectral radius of $P_{\mathcal{K}}$ is the same as that of

$$Q_{\mathcal{K}} = \begin{pmatrix} S_0 & \sqrt{S_0 S_1} & \cdots & \sqrt{S_0 S_{\ell-1}} & \sqrt{S_0 S_\ell} \\ \sqrt{S_0 S_1} & S_1 & \cdots & \sqrt{S_1 S_{\ell-1}} & 0 \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ \sqrt{S_0 S_{\ell-1}} & \sqrt{S_1 S_{\ell-1}} & \cdots & 0 & 0 \\ \sqrt{S_0 S_\ell} & 0 & \cdots & 0 & 0 \end{pmatrix},$$

where we write S_t instead of $S_t(\mathcal{K})$ for ease of notation.

Corollary 38 then gives $S_t(\mathcal{K}) = O(\alpha^t \log(n)|\mathcal{K}|) = o(\alpha^t \tau^{\ell/2})$, and the same calculation as in Proposition 17 yields:

$$\rho(Q_{\mathcal{K}}) = o(\alpha^{\ell/2}\tau^{\ell/2}) = o(\mu_2^\ell), \tag{43}$$

and the theorem follows.

■

Appendix F. Proof of Theorem 12

In order to prove Theorem 12, we need to show that the controls in the proof of Theorem 11 are actually sharp. We begin with the following lemma, which comes from the fact that ℓ-neighbourhoods of the vertices of G are roughly of the same size:

Lemma 39 *Assume that $\gamma = \Theta(\tau^{\ell/2})$. Then there exists a set of vertices $\mathcal{K}$ of size γ such that:*

$$S_\ell(\mathcal{K}) = \Omega(\alpha^\ell \cdot \gamma). \tag{44}$$

Proof Let $\varepsilon > 0$ to be determined later, S be the set consisting of the $n^{1-\varepsilon}$ vertices i with the largest values $S_\ell(i)$; we first show that, for all $i \in S$

$$S_\ell(i) = \Theta(\alpha^\ell). \tag{45}$$

Indeed, from Lemma 36, we have the folowing inequalities:

$$Kn\alpha^{2\ell} \leq \sum_{i=1}^n S_\ell(i)^2 \leq n \min_{i \in S} S_l(i)^2 + |S|(C\log(n)\alpha^\ell)^2, \tag{46}$$

and the second term is negligible before the two others, which implies (45).

We then build a set $\mathcal{K}$ of size γ as follows: begin with any member of S, and at each step add a vertex x such that $d(x, \mathcal{K}) > 2\ell$. This is possible as long as the 2ℓ-neighbourhood of $\mathcal{K}$ does not cover S, i.e. as long as:

$$\gamma \cdot C\log(n)\alpha^{2\ell} < n^{1-\varepsilon}. \tag{47}$$

But the LHS of this inequality is bounded by $C\log(n)n^{3/4}$, so this condition is satisfied as long as $\varepsilon < 1/4$.

By this construction, the vertices of $\mathcal{K}$ have ℓ-neighbourhoods that are pairwise disjoint, so by equation (39) we have:

$$S_\ell(\mathcal{K}) = \Omega(\alpha^\ell \times \gamma). \tag{48}$$

■

Consider now the vector v such that:

$$v_i = \begin{cases} \gamma^{-1/2} & \text{if } i \in \mathcal{K} \\ S_\ell(\mathcal{K})^{-1/2} & \text{if } d(i,\mathcal{K}) = \ell \\ 0 & \text{otherwise .} \end{cases} \tag{49}$$

The aim is to show the following equalities:

$$\frac{v^\top D^{(\ell)} v}{\|v\|^2} = \Omega(\mu_2^\ell) \quad \text{and} \quad \langle v, B^{(\ell)} \chi_k \rangle = o(\|v\| \|B^{(\ell)} \chi_k\|) \ \forall k \in [r_0] \tag{50}$$

Indeed, Theorem 12 will then follow from a simple application of Courant-Fisher's Theorem.

Proof (of Eq. (50). We notice that $\|v\|^2 = 2$; furthermore:

$$\begin{aligned} v^\top D^{(\ell)} v &= \sum_{i,j} v_i D_{ij}^{(\ell)} v_j \\ &\geq 2 \sum_{i \in S_\ell(\mathcal{K})} \sum_{j \in \mathcal{K}} v_i v_j \\ &= 2\gamma S_\ell(\mathcal{K}) \gamma^{-1/2} S_\ell(\mathcal{K})^{-1/2} \\ &= 2\sqrt{\gamma S_\ell(\mathcal{K})} \\ &= \Omega(\mu_2^\ell), \end{aligned}$$

which proves the first inequality.

It remains then to prove that v is asymptotically orthogonal to $B^{(\ell)}\chi_k$ for $k \in [r_0]$: noticing that $v_i \leq 1$ for all i and $\|v\|_0 = \gamma + S_\ell(\mathcal{K})$, we find, using Corollary 38:

$$\begin{aligned} \langle v, B^{(\ell)} \chi_k \rangle &\leq (\gamma + S_\ell(\mathcal{K})) \cdot \|B^{(\ell)} \chi_k\|_\infty \\ &\leq (\gamma + S_\ell(\mathcal{K})) \cdot \widetilde{O}(\alpha^\ell) \\ &= \widetilde{O}(\gamma \alpha^{2\ell}) \\ &= o(\sqrt{n} \mu_k^\ell) \quad \text{since } \kappa < 1/4 \\ &= o(\|v\| \|B^{(\ell)} \chi_k\|), \end{aligned}$$

where we used part (ii) of proposition 26 to bound $\|B^{(\ell)} \chi_k\|_\infty$. ∎