Non-backtracking spectra of
weighted inhomogeneous random graphs

Ludovic Stephan and Laurent Massoulié

Abstract. We study a model of random graphs where each edge is drawn independently (but not necessarily identically distributed) from the others, and then assigned a random weight. When the mean degree of such a graph is low, it is known that the spectrum of the adjacency matrix A deviates significantly from that of its expected value $\mathbb{E}A$.

In contrast, we show that over a wide range of parameters the top eigenvalues of the non-backtracking matrix B – a matrix whose powers count the non-backtracking walks between two edges – are close to those of $\mathbb{E}A$, and all other eigenvalues are confined in a bulk with known radius. We also obtain a precise characterization of the scalar product between the eigenvectors of B and their deterministic counterparts derived from the model parameters.

This result has many applications, in domains ranging from (noisy) matrix completion to community detection, as well as matrix perturbation theory. In particular, we establish as a corollary that a result known as the Baik–Ben Arous–Péché phase transition, previously established only for rotationally invariant random matrices, holds more generally for matrices A as above under a mild concentration hypothesis.

1. Introduction

Let $P \in \mathcal{M}_n(\mathbb{R})$ be a symmetric $n \times n$ matrix with entries in $[0, 1]$, and W a (symmetric) weight matrix with independent random entries. We define the inhomogeneous undirected random graph $G = (V, E)$ associated with the couple (P, W) as follows: the vertex set is simply $V = [n]$, and each edge $\{u, v\}$ is present in E independently with probability P_{uv}, and holds weight W_{uv}.

The entrywise expected value and variance of the weighted adjacency matrix of G are

$$\mathbb{E}A = P \circ \mathbb{E}W, \quad \mathrm{Var}(A) := P \circ \mathbb{E}[W \circ W] - P \circ P \circ \mathbb{E}W \circ \mathbb{E}W, \qquad (1.1)$$

where $\circ$ denotes the Hadamard product. When the entries of P are small, the second term of $\mathrm{Var}(A)$ is negligible and the variance can be well approximated by the

---

entrywise second moment; we thus define

$$Q := P \circ \mathbb{E}W, \quad K := P \circ \mathbb{E}[W \circ W]. \tag{1.2}$$

A natural question, arising from matrix perturbation theory, is then as follows:

What is the relationship between the eigendecomposition of A and the one of Q?

Unfortunately, at least in the unweighted case, when the mean degree of G is low ($o(\log(n))$), it is known that the largest eigenvalues (and associated eigenvectors) of A are determined by the large degree vertices; see [9] for a complete description of this phenomenon. To extract meaningful information on the spectrum of Q, another matrix has shown better performance: the non-backtracking matrix, whose application to community detection has been studied in [14, 25].

Given a weighted graph G, we define its associated non-backtracking matrix B as follows: B is a $2|E| \times 2|E|$ matrix indexed by the oriented edges of G, whose coefficients are

$$B_{ef} = W_f \mathbf{1}\{e \to f\} = W_f \mathbf{1}\{e_2 = f_1\}\mathbf{1}\{e_1 \neq f_2\},$$

where $e = (e_1, e_2)$ and $f = (f_1, f_2)$. The above question rephrases in our setting as

What is the relationship between the eigendecomposition of B and the one of Q?

and the main focus of this article is to provide an answer as precise as possible to this problem. To this end, let

$$Q = \sum_{i=1}^{r} \mu_i \varphi_i \varphi_i^\top \quad \text{with } |\mu_1| \geq |\mu_2| \geq \cdots \geq |\mu_r|$$

be the eigendecomposition of Q, and $\rho = \rho(K)$ the largest eigenvalue (in absolute value) of K. Note that by definition, Q and K are symmetric and therefore all eigenvalues defined above are real.

We shall assume that there exists some deterministic bound L (possibly depending on n) such that $\max |W_{ij}| \leq L$. We can then state our main theorem, without detailing the needed hypotheses for now:

Theorem 1 (informal statement). *Assume the following conditions:*

(i) $r = n^{o(1)}$,

(ii) *the graph G is sparse enough,*

(iii) *the eigenvectors of Q are sufficiently delocalized.*

Let r_0 be the number of eigenvalues of Q whose absolute value is larger than both $\sqrt{\rho}$ and L:

$$\mu_k > \sqrt{\rho} \vee L \text{ for all } k \in [r_0], \quad \mu_{r_0+1} \leq \sqrt{\rho} \vee L. \tag{1.3}$$

Then, for $i \leq r_0$, the i-th largest eigenvalue of B is asymptotically (as n goes to infinity) equal to μ_i, and all the other eigenvalues of B are constrained in a circle of center 0 and radius $\max(\sqrt{\rho}, L)$. Further, if $i \leq r_0$ is such that μ_i is a sufficiently isolated eigenvalue of Q, then the eigenvector associated with the i-th eigenvalue of B is correlated to a lifted version of φ_i.

Next section consists in the detailed statement of this theorem (with precise hypotheses and bounds given).

2. Detailed setting and results

2.1. Notations

General notations. Throughout this paper, we use the following notations:

- for integer n, $[n]$ denotes the set $\{1, \ldots, n\}$.

- for $x \in \mathbb{R}^n$, we shall denote by x_i or $x(i)$ the i-th coordinate of x, whichever is most convenient. $\|x\|$ is the 2-norm of x, and $\|x\|_\infty$ the infinity norm of x.

- the operator norm of a matrix M is noted $\|M\|$; it is the maximal singular value of M. Its Frobenius norm is denoted $\|M\|_F$ and its infinity norm $\|M\|_\infty = \sup_{i,j} |M_{ij}|$.

- $\mathbf{1}$ denotes the all-one vector, and $\mathbf{1}\{\cdot\}$ is the indicator function of an event.

- the group of permutations on r elements is noted $\mathfrak{S}_r$.

- the max (resp. min) of two numbers a, b is noted $a \vee b$ (resp. $a \wedge b$).

- the letter c denotes any absolute constant, whose value should be assumed to be the maximum of any such constant encountered so far. To improve the readability of our computations, we use numbered constants c_i during proofs.

Graph theoretic notations. For a graph $g = (V, E)$, let $\vec{E}$ be the set of oriented edges in E, and

$$\vec{E}(V) = \{(u, v) \mid u \neq v \in V\}$$

be the set of all directed edges of the *complete graph* on V. If t is an integer, $g = (V, E)$ is a graph and $x \in V$, then the *ball* $(g, x)_t$ is the subgraph induced by all edges at distance at most t from x, and $\partial(g, x)_t$ is the boundary of the ball, i.e. the set of vertices at distance exactly t from x. Finally, the set of all non-backtracking paths of length t starting with x will be denoted $\mathcal{P}_g(x, t)$.

Non-backtracking matrix. Since we are interested in the spectrum of the non-backtracking matrix B, we need to be able to translate "vertex" quantities such as the

vectors φ_i into "edge" quantities. Recall that $V = [n]$, and identify $\vec{E}$ with the set $[2m]$; we define the $2m \times n$ *start* and *terminal* matrices S and T as

$$\forall e \in \vec{E}, \; i \in [n], \quad S_{ei} = \mathbf{1}\{e_1 = i\}, \quad T_{ei} = \mathbf{1}\{e_2 = i\}. \tag{2.1}$$

For a vector $\phi \in \mathbb{R}^n$, this implies that $[T\phi](e) = \phi(e_2)$ for every edge $e \in \vec{E}$. We then define the "lifted" eigenvectors $\chi_i = T\varphi_i$ for $i \in [r]$.

We also define the *reverse* operator J such that $Je = \bar{e} := (e_2, e_1)$, and the diagonal matrix D_W such that $D_W(e, e) = W_e$; from the definition of B and symmetry of W it is straightforward to see that $JD_W = D_W J$ and for all $t \geq 0$,

$$JD_W B^t = (B^*)^t D_W J, \tag{2.2}$$

which is known in mathematical physics as parity-time invariance. For any vector $x \in \mathbb{R}^{\vec{E}}$, we denote the vector Jx by $\check{x}$.

Building upon the sketch in the introduction, we now expand on the model definition. Recall that the *expectation* and *variance* matrices were defined as

$$Q = P \circ \mathbb{E}W, \quad K = P \circ \mathbb{E}[W \circ W].$$

2.2. Defining the convergence parameters

In full generality, with no assumptions on P and W, we do not expect meaningful results to hold; however, we are still able to provide interesting properties on a large class of matrices. We define in the following the parameters that will govern the convergence behavior:

(i) the *rank*

$$r = \max\big(\mathrm{rank}(Q), \sqrt{\mathrm{rank}(K)}\big);$$

note that in most practical applications (such as the unweighted case), we shall have $r = \mathrm{rank}(Q)$, but we also treat cases where $r \gg \mathrm{rank}(Q)$.

(ii) the sparsity parameter

$$d = n \max_{i,j \in [n]} P_{ij};$$

(iii) the eigenvector delocalization parameter

$$b = \sqrt{n} \max_{i \leq \mathrm{rank}(Q)} \|\varphi_i\|_\infty;$$

(iv) the signal-to-noise ratio

$$\tau = \max_{\mu_i^2 > \mu_1} \frac{\mu_1}{\mu_i^2};$$

(v) and finally the almost sure probability bound

$$\|W\|_\infty = L;$$

our results hold trivially whenever $L = +\infty$ so we shall restrict ourselves to the case where L is finite, and the W_{ij} are almost surely bounded. While Theorem 2 below requires an almost sure bound, techniques for dealing with high probability bounds are discussed in Theorem 6.

The average degree of a vertex i will be noted by

$$d_i = \sum_{j \in [n]} P_{ij} \leq d,$$

which corresponds to the entries of the vector $P\mathbf{1}$. To ensure that G is connected enough for spectral properties to hold, we make the (common) assumption that $d_i \geq 1$ for all $i \in [n]$. The entries of $K\mathbf{1}$ can be viewed as an extension of the average degrees in the weighted case (see [5] or [10] for examples), and for the same reason as above we require that $K\mathbf{1}$ is bounded away from zero by a constant.

2.3. Main theorem

In the following, $G = G(P, W)$ is the random graph defined in the introduction, B is the non-backtracking matrix associated with G, and $|\lambda_1| \geq \cdots \geq |\lambda_{2m}|$ are its eigenvalues.

In its most general form, our main result is as follows:

Theorem 2. *Let $n \geq 0$ and (P, W) be a couple of $n \times n$ matrices defining a random graph G. Define $\rho = \rho(K)$, r_0 as in (1.3), r, b, d, τ, L as in Section 2.2, and $\tilde{L} = L/\mu_1$.*
  Let
$$\ell = \frac{1 - \epsilon}{16} \frac{\log(n)}{\log(d)},$$

for arbitrary $\epsilon > 0$. There exist numbers n_0 and C_0, all depending on n and the convergence parameters, such that the following holds:

(i) *C_0 is smaller than*
$$c \left(\frac{rbd\tilde{L}\log(n)}{1 - \tau} \right)^{25},$$

  and n_0 is smaller than

$$\exp\left(c \frac{\max\{\log(r), \log(b), \log(d)^2, \log(\tilde{L}), \log(\log(n))\}}{\log(\tau^{-1})} \right).$$

(ii) *If $n \geq n_0$, define*

$$\sigma := C_0 \mu_1 \tau^{\ell/2}. \tag{2.3}$$

Then the following holds with probability at least $1 - c/\log(n)$, there exists a permutation s of $[r_0]$ such that

$$\max_{i \in [r_0]} |\lambda_i - \mu_{s(i)}| \leq \sigma, \tag{2.4}$$

and all the remaining eigenvalues of B are less than $C_0^{1/\ell}(\sqrt{\rho} \vee L)$.

(iii) *For any $i \in [r_0]$, if*

$$\delta_i := \min_{j \neq s(i)} |\mu_{s(i)} - \mu_j| \geq 2\sigma, \tag{2.5}$$

then there exists a normed eigenvector ξ associated with λ_i such that

$$\langle \xi, \xi_i \rangle \geq \sqrt{1 - rd^2 \widetilde{L}^2 \frac{\rho}{\mu_i^2}} + O\left(\frac{\sigma}{\delta_i - \sigma}\right), \quad \text{where } \xi_i = \frac{T\varphi_i}{\|T\varphi_i\|}.$$

In order to get an applicable and useful result, we need $n \geq n_0$ when n is sufficiently large, and $C_0^{1/\ell}$ goes to 1 as n goes to infinity. Both conditions are verified in particular when

$$1 - \tau = \Omega(1), \quad r, b = n^{o(1)}, \quad \log(d)^2 = o(\log(n)).$$

By definition of $\widetilde{L}$, whenever $\widetilde{L} > 1$ we have $\mu_1 < L$, and thus $r_0 = 0$. We can therefore safely assume $\widetilde{L} \leq 1$ in applications and not focus on any bound for L.

The proof of this theorem follows the same method as in many spectral proofs, from [28] to more recent papers such as [13]. It consists of the following:

- show that the neighborhood of any vertex v is close to a suitably defined random tree,

- study a family of graph functionals that give rise to approximate eigenvectors of the random tree,

- use a concentration argument to transpose those tree eigenvectors to pseudo-eigenvectors of the non-backtracking matrix,

- bound the remaining eigenvalues using a variant of the trace method in [20],

- conclude by a matrix perturbation argument.

A large portion of the remainder of this paper is dedicated to implementing this method; however, we first provide several applications of our result to the fields of random matrix theory and random graph theory.

3. Applications

3.1. Phase transition in random graphs

Matrix perturbation theory focuses on finding the eigenvalues and eigenvectors of matrices of the form $X + \Delta$, where X is a known matrix and Δ is a perturbation assumed "small" in a sense. Celebrated results in this field include the Bauer–Fike theorem [8] for asymmetric matrices, and the Weyl [34] and Davis–Kahan [35] theorems for symmetric ones; incidentally the present paper makes use of those results in its proofs. Finding sharp general theorems without additional assumptions is known to be hard, since the eigenvalues and eigenvectors depend on the interactions between the eigenspaces of X and Δ.

In the last two decades, growing attention has been paid to problems of the following form: finding the eigenvectors of $X_n + P_n$ (or, in its multiplicative form, $X_n(I_n + P_n)$), where P_n is an $n \times n$ matrix with low rank $r \ll n$ (usually fixed) and known eigenvalues, and X_n is a random matrix with known distribution. Examples of this setting are the spiked covariance model [6, 23] and additive perturbations of Wigner matrices [16, 19, 30]. A more systematic study has been performed in [10, 12] on orthogonally invariant random matrices.

A staple of those results is the existence of a so-called *BBP phase transition* (named after Baik–Ben Arous–Péché, from the seminal article [6]): in the limit $n \to \infty$, each eigenvalue of P_n that is above a certain threshold gets reflected (albeit perturbed) in the spectrum of $X_n + P_n$, with the associated eigenvector correlated to the one of P_n.

Phase transition for the adjacency matrix. The adjacency matrix A of our random graph G can be viewed as a perturbation model by writing

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

The term $\operatorname{diag}(Q)$ being negligible with respect to the others, we can see A as the sum of a deterministic low-rank matrix and a random noise matrix with i.i.d. centered entries. Further, the entrywise variance of A is equal (up to a negligible term) to K, so the parameter ρ can be seen as an equivalent to the variance in the Wigner model. We thus expect, whenever $\sqrt{\rho} \gg L$ (so that $\sqrt{\rho}$ is the actual threshold in Theorem 2), to find a phase transition akin to the one in [11]; and indeed the following theorem holds:

Theorem 3. *Let (P, W) be a matrix couple of size $n \times n$ and r, b, d, τ, L as above. Assume further that:*

(i) *the Perron–Frobenius eigenvector of K is $\mathbf{1}$; that is $K\mathbf{1} = \rho\mathbf{1}$,*

(ii) *the above eigenvector equation concentrates, i.e. with high probability there exists $\varepsilon \leq 1/2$ such that for all $i \in [n]$,*

$$\left| \sum_{j \sim i} W_{ij}^2 - \rho \right| \leq \varepsilon \rho. \tag{3.1}$$

Then, if $i \in [r_0]$ is such that $\mu_i^2 \geq 2L^2$, there exists an eigenvalue ν_i of A that verifies

$$\nu_i = \mu_i + \frac{\rho}{\mu_i} + \frac{\rho}{\mu_i} \cdot O\left(\frac{L}{\mu_i} + \frac{L^2}{\mu_i^2} + \varepsilon \right). \tag{3.2}$$

Further, if the mean degree d_j for all j is equal to $d_0 > 1$, and i is such that $\delta_i \geq 2\sigma$ (with σ and δ_i defined in (2.3) and (2.5)), then there exists a normed eigenvector ζ of A with corresponding eigenvalue ν_i such that

$$\langle \zeta, \varphi_i \rangle = \sqrt{1 - \frac{\rho}{\mu_i^2}} + O\left[\frac{1}{\delta_i - \sigma} \left(\frac{L\rho}{\mu_i^2} + \frac{L^2 \rho}{\mu_i^3} + \varepsilon \frac{\rho}{\mu_i} \right) \right]. \tag{3.3}$$

Whenever $\rho \gg L^2$, and ε goes to zero as $n \to \infty$, then the condition $\mu_i^2 \geq 2L^2$ is always verified and the $O(\cdot)$ term in (3.2) vanishes, and the obtained expansion is therefore asymptotically correct. The presence of δ_i renders a similar result on the scalar product harder to obtain; however, assuming $\delta_i = \Theta(\sqrt{\rho})$ (that is, the eigenvalues of Q are somewhat regularly spaced) implies similarly that the $O(\cdot)$ term in (3.3) vanishes.

The obtained expression for ν_i, as well as the scalar product expansion, are identical to the ones in [11], for low-rank additive perturbations of Gaussian Wigner matrices. Our result is thus a direct extension of [11], for a larger class of matrices upon a sparsity and concentration condition. Such an extension is not unexpected, in view of results concerning the universality of the semicircle law for Bernoulli random matrices, such as [17].

An especially interesting particular case of Theorem 3 is the unweighted random graph setting, where $W_{ij} = 1$ for all i, j. In this case, we have $K = P$ so the eigenvector equation $K\mathbf{1} = \rho\mathbf{1}$ is equivalent to all the average degrees being equal, i.e. $d_i = d_0 = \rho$ for $i \in [n]$. It is a well-known fact (see for example [18]) that for unweighted random graphs the degree concentration property holds with $\varepsilon = 2\sqrt{\log(n)/d_0}$. A slight modification of the proof of Theorem 3 further removes several error terms, and the following corollary ensues:

Corollary 1. *Let P be a $n \times n$ matrix and r, b, d, τ as above, with $W = \mathbf{1}^*\mathbf{1}$. Assume further that for all $i \in [n]$,*

$$\sum_{j \in [n]} P_{ij} = d_0 > 16 \log(n).$$

Then for all $i \in [r_0]$, there exists an eigenvalue v_i of A that verifies

$$v_i = \mu_i + \frac{d_0}{\mu_i} + O\left(\sqrt{\frac{\log(n)}{d_0}} \frac{d_0}{\mu_i}\right),$$

and if i is such that $\delta_i > 2\sigma$, there exists a normed eigenvector of A such that

$$\langle \zeta, \varphi_i \rangle = \sqrt{1 - \frac{d_0}{\mu_i^2}} + O\left(\frac{1}{\delta_i - \sigma} \sqrt{\frac{\log(n)}{d_0}} \frac{d_0}{\mu_i}\right).$$

In particular, we have

$$v_1 = d_0 + 1 + O\left(\sqrt{\frac{\log(n)}{d_0}}\right).$$

This is an improvement on the results of [10], which only give $v_i = \mu_i + O(\sqrt{d_0})$. The condition $d_0 > 16 \log(n)$ ensures that the degrees of G concentrate. Since our result is really only meaningful whenever $d_0 \gg \log(n)$, so that the error term is negligible before d_0/μ_i, we do not perform the same detailed analysis as in [5]. However, a more precise phase transition around $d_0 \asymp \log(n)$ is not excluded.

Theorem 3 is derived from Theorem 2 through an adaptation of the Ihara–Bass formula [7], obtained by expanding arguments from [9, 33]:

Proposition 1. *Let x be an eigenvector of the matrix B with associated eigenvalue λ, such that $\lambda^2 \neq W_{ij}^2$ for every i, j. Define the weighted adjacency matrix $\tilde{A}(\lambda)$ and the diagonal degree matrix $\tilde{D}(\lambda)$ by*

$$\tilde{A}(\lambda)_{ij} = \mathbf{1}\{i \sim j\} \frac{\lambda W_{ij}}{\lambda^2 - W_{ij}^2}, \quad \tilde{D}(\lambda)_{ii} = \sum_{j \sim i} \frac{W_{ij}^2}{\lambda^2 - W_{ij}^2}.$$

Then the vector $y = S^* D_W x$, where S is defined as in (2.1), is a null vector of the laplacian matrix

$$\Delta(\lambda) = I - \tilde{A}(\lambda) + \tilde{D}(\lambda).$$

The details and computations are left to the appendix.

3.2. Community detection in random networks

Community detection is a clustering problem that aims to identify large subgroups (or communities) with similar characteristics inside a large population, with the only data available being the pairwise interactions between individuals. Starting from its introductory paper [22], the stochastic block model has been a popular generative

model for algorithm design; it consists of a random graph G where vertices are partitioned randomly in communities, and edges are present independently with probability depending only on the community membership of their endpoints. Popular algorithms for recovering communities include semi-definite programming methods [29], belief propagation [4], and spectral methods [26,28]; a comprehensive review of algorithms and results can be found in [1].

Unlabeled stochastic block model. In a general form, we can define the stochastic block model SBM(n, r, θ, M), where $\theta \in [r]^n$ and $M \in [0, 1]^{r \times r}$ as follows:

- the vertex set is $V = [n]$,

- each vertex $i \in [n]$ has a community label θ_i in $[r]$,

- for any pair of vertices (i, j), an edge is present between i and j independently from the others with probability $M_{\theta_i \theta_j}$.

It is common to assume $M = \frac{\alpha}{n} M_0$, where M_0 does not depend on n and α is a scaling parameter. It is easy to see that up to diagonal terms, the expected adjacency matrix has the form

$$P = \Theta M \Theta^*,$$

where Θ is a $n \times r$ matrix such that $\Theta_{ij} = 1$ if $\theta_i = j$, and 0 otherwise. We shall assume that for any $k \in [r]$,

$$\frac{\#\{i \in [n] \mid \theta_i = k\}}{n} = \pi_k > 0, \tag{3.4}$$

where π is a deterministic probability vector. Let $\mu_1 \geq \cdots \geq |\mu_r|$ the eigenvalues of $\mathrm{diag}(\pi) M_0$, with α chosen such that $|\mu_1| = 1$, and $\phi_1, \ldots, \phi_r$ the associated eigenvectors. Then the non-zero eigenvalues of P are easily found to be the $\alpha \mu_i$, with associated eigenvectors $\Theta \phi_i$.

A common assumption is that each vertex type has the same average degree, i.e.

$$P\mathbf{1} = \alpha \mathbf{1},$$

otherwise a simple clustering based on vertex degree correlates with the underlying communities. Making this additional assumption, the following theorem holds:

Theorem 4. *Assume that r is constant, and $\alpha = n^{o(1)}$. Let r_0 be defined as follows:*

- *if $\alpha \geq 1$ is constant, r_0 is the only integer in $[r]$ such that*

$$\alpha \mu_k^2 > 1 \quad \text{for } i \in r_0, \quad \alpha \mu_{r_0+1}^2 \leq 1.$$

- *if $\alpha = \omega(1)$, then $r_0 = r$.*

Then, for any n larger than an absolute constant and all $i \in [r_0]$, one has

$$|\lambda_i - \mu_i| \leq c(\alpha \log(n))^a (\alpha \mu_{r_0})^{-\kappa \log_\alpha(n)} := \sigma$$

for some positive constants c, a, κ, and all other eigenvalues of B are confined in a circle with radius $(1 + o(1)) \sqrt{\alpha}$. Further, if μ_i is an isolated eigenvalue of $\mathrm{diag}(\pi) M_0$, then there exists an eigenvector ξ of the non-backtracking matrix B associated with λ_i such that

$$\langle \xi, \xi_i \rangle \geq \sqrt{1 - \frac{1}{\alpha \mu_i^2}} + O(\sigma'), \quad where \; \xi_i = \frac{T \Theta \phi_i}{\|T \Theta \phi_i\|}.$$

This theorem is essentially a corollary of Theorem 2, with some simplifications due to $Q = K = P$ and $P\mathbf{1} = \alpha \mathbf{1}$; the error bound σ is the same as in the main theorem. It is a direct generalization of Theorem 4 in [14], for a diverging degree sequence; further, the property $\langle \xi, \xi_i \rangle = 1 - o(1)$ as soon as $\alpha \gg 1$ suggests that a clustering algorithm such as k-means performed on the eigenvectors of B recovers all but a vanishing fraction of the community memberships in this regime, which would provide an alternative to the Sphere-comparison algorithm presented in [3].

Conjecture 1. *In the SBM defined as above, as soon as $\alpha = \omega(1)$, running an approximate k-means algorithm on the top r eigenvectors of B allows to recover the community memberships of every vertex but a vanishing fraction as $n \to \infty$.*

Proving this conjecture would require a more careful eigenspace analysis for eigenvalues with multiplicity more than one, such as the one performed in [32], as well as an error bound on the clustering step similar to the one in [26].

Remark. When the expected degrees of each vertex type is not the same, an analogous version of Theorem 4 holds. The main difference in this case is that the scalar product $\langle \xi, \xi_i \rangle$ has a less elegant asymptotic expansion.

Since the lead eigenvector of P is now non-constant, the condition for reconstruction is simply $\alpha > 1$ (or $r_0 \geq 1$). In particular, this is true as long as the average degree of the graph is above one; indeed by the Courant–Fisher principle

$$\alpha > \frac{\mathbf{1}^\top P \mathbf{1}}{n}.$$

Labeled block models. In real-world networks, pairwise interactions often carry more information than just a binary one. A popular variant of the stochastic block model is thus a model with added edge labels, as follows: let $\mathcal{L}$ be a label space, and consider a SBM drawn under the model described above. We assign to an edge (i, j) a label $L_{ij} \in \mathcal{L}$, drawn independently from a distribution $\mathbb{P}_{\theta_i \theta_j}$. Such classes of models have been investigated in full generality in [21, 27], and a variant with the underlying graph being an Erdős–Rényi model in [31].

We shall focus here on the symmetric two-community SBM, with

$$\pi = \left(\frac{1}{2}, \frac{1}{2}\right), \quad M = \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \quad \mathbb{P}_{11} = \mathbb{P}_{22} = \mathbb{P}, \quad \mathbb{P}_{12} = \mathbb{P}_{21} = \mathbb{Q}, \quad (3.5)$$

and assume that both measures are absolutely continuous with respect to another measure m (note that we can take $m = \mathbb{P} + \mathbb{Q}$), with Radon–Nikodym derivatives f and g. Let $w: \mathcal{L} \to \mathbb{R}$ a bounded weight function, such that $w(\ell) \leq L$ for any $\ell \in \mathcal{L}$; and define the weight matrix $W_{ij} = w(L_{ij})$ and the associated weighted non-backtracking matrix B. Then, an application of Theorem 2 yields the following result:

Theorem 5. *Define the parameter τ by*

$$\tau = 2\frac{(a\mathbb{E}_{\mathbb{P}}[w^2] + b\mathbb{E}_{\mathbb{Q}}[w^2]) \vee L}{(a\mathbb{E}_{\mathbb{P}}[w] - b\mathbb{E}_{\mathbb{Q}}[w])^2}.$$

Then, whenever $\tau < 1$, let ξ be a normed eigenvector corresponding to the second eigenvalue of B. There exists a parameter

$$\sigma \leq (a\log(n))^{25}\tau^{\kappa\log_a(n)}$$

for some constant κ such that

$$\langle \xi, \xi_0 \rangle = \sqrt{1-\tau} + O(\sigma), \quad \text{where } \xi_0 = \frac{\Theta\binom{1}{-1}}{\sqrt{n}}.$$

Whenever this result holds, a proof identical to the one in [28] implies that recovering a positive fraction of the community memberships is possible.

In order to maximize the region in which reconstruction is possible, we need to choose the weights $w(\ell)$ such that τ is minimized. This optimization step is performed in the appendix, and leads to the following:

Proposition 2. *Define the weight function w and signal-to-noise ratio β as*

$$w(\ell) = \frac{af(\ell) - bg(\ell)}{af(\ell) + bg(\ell)}, \quad \beta = \frac{1}{2}\int \frac{(af - bg)^2}{af + bg}\,\mathrm{d}m, \quad (3.6)$$

where a, f, b, g and m are defined in equation (3.5) and below. Then, whenever $\beta > 1$, a spectral algorithm based on the matrix B is able to recover a positive fraction of the community memberships when $n \to \infty$.

This settles a conjecture of [21], generalizing the setting from finite to arbitrary label space. Whenever we allow for a higher number of communities, as well as arbitrary choices for the connectivity matrix Q and distributions $\mathbb{P}_{ij}$, the problem proves

to be harder; an analog to Theorem 5 does hold, but the optimization problem required to minimize the ratio τ looks to be untractable. In the symmetric SBM case, where

$$\pi = \frac{1}{k}, \quad M = a\mathbf{1}\{i = j\} + b\mathbf{1}\{i \neq j\}, \quad \mathbb{P}_{ij} = \mathbb{P}\mathbf{1}\{i = j\} + \mathbb{Q}\mathbf{1}\{i \neq j\},$$

we make the following conjecture:

Conjecture 2. *In the labeled symmetric SBM, partial reconstruction is possible as soon as $\beta > 1$, where*

$$\beta = \frac{1}{k} \int \frac{(af - bg)^2}{af + (k-1)bg} \, dm,$$

and a spectral algorithm based on the non-backtracking matrix with weight function

$$w(\ell) = \frac{af(\ell) - bg(\ell)}{af(\ell) + (k-1)bg(\ell)}$$

recovers a positive fraction of the community memberships in polynomial time.

As with Theorem 4, whenever the mean degree α of the graph grows to infinity, we have $\langle \xi, \xi_0 \rangle = 1 - o(1)$, which brings us our second conjecture:

Conjecture 3. *If we have $a = \alpha a_0$, $b = \alpha b_0$ with $\alpha = \omega(1)$, a_0, b_0 fixed, then as $n \to \infty$ a clustering algorithm based on the second eigenvector of the weighted non-backtracking matrix B with the weight function defined in (3.6) recovers all but a vanishing fraction of the community memberships.*

As a final remark, note that the optimal weight function assumes perfect knowledge of all model parameters, especially the exact label distribution for each community pair. However, in some cases, this weight function is a rescaling of a more agnostic one; as an example, in the censored block model [2] we find that $w(\ell) = c\ell$ (with $\ell = \pm 1$), and thus the spectral algorithm mentioned here is the same as in [31].

3.3. Extension to gaussian weights

In the form presented in Theorem 2, our result is only meaningful with almost surely bounded random variables (i.e. with $L < \infty$). With a more careful analysis of the error bounds, this can be extended to

$$L = \sup_{i,j \in [n]} \sup_k \mathbb{E}\big[W_{ij}^k\big]^{1/k}; \tag{3.7}$$

however, we determined the class of distributions satisfying (3.7) was not different enough from the bounded case to warrant increasing the complexity of the proof.

To the contrary, the setting where the W_{ij} are gaussian random variables is of independent interest; it can be seen as a special case of noisy matrix completion as described in [15, 24]. In this case, the moment condition of (3.7) is far from satisfied, and at least at first glance our proof cannot be adapted readily. Still, we show the following:

Theorem 6. *Assume that the $W_{ij} \sim \mathcal{N}(m_{ij}, s_{ij}^2)$ are independent Gaussian random variables, and let*

$$m = \sum_{i,j} m_{ij}, \quad s = \sup_{i,j} s_{ij}.$$

Then the conclusions of Theorem 2 apply with

$$L = m + 2s\sqrt{\log(n)}.$$

The loss of a $\sqrt{\log(n)}$ factor comes from the use of a concentration bound for the W_{ij}; details can be found in the appendix.

To the best of our knowledge, there is not much literature to compare with on the topic of eigenvalue reconstruction for noisy matrix completion, the works cited above being focused on reconstructing the whole matrix. However, results on gaussian matrix perturbation such as [11] seem to indicate that the $\sqrt{\log(n)}$ factor is superfluous and can be improved upon with other methods.

4. A Bauer–Fike type bound for almost orthogonal diagonalization

One important tool in tying together the local analysis of G is a matrix perturbation theorem, derived from the Bauer–Fike theorem. It mostly consists in a simplification and adaptation of Theorem 8.2 in [13], tailored to our needs. We begin by recalling the original Bauer–Fike theorem:

Theorem 7 (Bauer–Fike theorem [8]). *Let D be a diagonalizable matrix, such that $D = V^{-1}\Lambda V$ for some invertible matrix V and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. Let E be any matrix of size $n \times n$. Then, any eigenvalue μ of $D + E$ satisfies*

$$|\mu - \lambda_i| \leq \|E\|\kappa(V), \tag{4.1}$$

for some $i \in [n]$, where $\kappa(V) = \|V\|\|V^{-1}\|$ is the condition number of V.

Let R be the right-hand side of (4.1), and $C_i := \mathcal{B}(\lambda_i, R)$ the ball centered at λ_i with radius R (in $\mathbb{C}$). Let $\mathcal{I} \subseteq [n]$ be a set of indices such that

$$\left(\bigcup_{i \in \mathcal{I}} C_i\right) \cap \left(\bigcup_{i \notin \mathcal{I}} C_i\right) = \emptyset.$$

Then the number of eigenvalues of $D + E$ in $\bigcup_{i \in \mathcal{I}} C_i$ is exactly $|\mathcal{I}|$.

4.1. A custom perturbation lemma for almost diagonalizable matrices

Building on this theorem, we now expose this section's first result. Let $U = (u_1, \ldots, u_r)$ and $V = (v_1, \ldots, v_r)$ be $n \times r$ matrices; our nearly diagonalizable matrix shall be $S = U\Sigma V^*$ with $\Sigma = \mathrm{diag}(\theta_1, \ldots, \theta_r)$. We shall assume that the θ_i are in decreasing order of modulus:

$$|\theta_r| \leq |\theta_{r-1}| \leq \cdots \leq |\theta_1| = 1.$$

Now, let A be a $n \times n$ matrix, not necessarily diagonalizable. The assumptions needed for our results are as follows:

(i)   For some small constant $\varepsilon > 0$,

$$\|A - S\| \leq \varepsilon.$$

(ii)   The matrices U and V are well-conditioned: both U^*U and V^*V are non-singular, and there exist two constants $\alpha, \beta > 1$ such that

$$\|U^*U\| \leq \alpha, \qquad \|V^*V\| \leq \alpha,$$
$$\|(U^*U)^{-1}\| \leq \beta, \quad \|(V^*V)^{-1}\| \leq \beta.$$

(iii) There exists another constant $0 < \delta < 1$ such that

$$\|U^*V - I_r\|_\infty \leq \delta.$$

(iv)   The θ_i are well-separated from 0, in the sense that

$$|\theta_r| > 2\sigma, \tag{4.2}$$

where an exact expression for σ will be given over the course of the proof.

Then the following result, whose statement and proof (regarding the eigenvalue perturbation) are adapted from [13], holds:

Theorem 8. *Let A be a matrix satisfying assumptions* (i)–(iv) *above, and let $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_r|$ be the r eigenvalues of A with largest modulus. There exists a permutation π such that for all $i \in [r]$,*

$$|\lambda_{\pi(i)} - \theta_i| \leq r \times \sigma,$$

and the other $n - r$ eigenvalues of A all have modulus at most σ. Additionally, if i is such that

$$B(\theta_i, \sigma) \cap \left(\bigcup_{j \neq i} B(\theta_j, \sigma) \right) = \emptyset, \tag{4.3}$$

then there exists a normed eigenvector ξ associated with $\lambda_{\pi(i)}$ such that

$$\left\| \xi - \frac{u_i}{\|u_i\|} \right\| \le \frac{3\sigma}{\delta_i - \sigma},$$

where δ_i is the minimum distance from θ_i to another eigenvalue:

$$\delta_i = \min_{j \ne i} |\theta_j - \theta_i| \ge 2\sigma.$$

Proof. We begin with defining an alternative matrix $\bar{U}$ such that $\bar{U}^* V = I_r$. Let H_i be the subspace of $\mathbb{R}^n$ such that

$$H_i = \text{vect}(v_j \mid j \ne i),$$

and consider the vectors $\tilde{u}_i$ and $\bar{u}_i$ defined as

$$\tilde{u}_i = u_i - P_{H_i}(u_i), \quad \bar{u}_i = \frac{\tilde{u}_i}{\langle \tilde{u}_i, v_i \rangle}$$

with P_{H_i} the projection on H_i, and $\tilde{U}$, $\bar{U}$ the associated $n \times r$ matrices. Then it is straightforward to see that

$$\langle \bar{u}_i, v_i \rangle = 1, \quad \langle \bar{u}_i, v_j \rangle = 0$$

for all $j \ne i$, which shows that $\bar{U}^* V = I_r$. Now, if we let V_i be the matrix V with the i-th column and row deleted,

$$P_{H_i} = V_i (V_i^* V_i)^{-1} V_i^*,$$

and

$$\|V_i^* u_i\|^2 = \sum_{j \ne i} \langle v_j, u_i \rangle^2 \le r\delta^2,$$

hence we can compute $\|u_i - \tilde{u}_i\|$:

$$\|u_i - \tilde{u}_i\| = \|P_{H_i}(u_i)\| \le \|V_i\| \|(V_i^* V_i)^{-1}\| \|V_i^* u_i\|,$$

and by the interlacing theorem $\|V_i\| \le \sqrt{\alpha}$ and $\|(V_i^* V_i)^{-1}\| \le \beta$ since V_i is a principal submatrix of V. Using the fact that $\|M\| \le \|M\|_F$ for any matrix M, we find

$$\|U - \tilde{U}\| \le r^2 \sqrt{\alpha} \beta \delta.$$

For the second part, note that by the Cauchy–Schwarz inequality,

$$\begin{aligned}
|\langle \tilde{u}_i, v_i \rangle - 1| &\le |\langle u_i, v_i \rangle - 1| + \|u_i - \tilde{u}_i\| \cdot \|v_i\| \\
&\le \delta(1 + r\alpha\beta),
\end{aligned}$$

with the (generous) inequality $\|v_i\| \leq \|V\|$ used in the last line. Whenever δ is small enough, we can use the inequality $|(1-t)^{-1} - 1| \leq 2t$ which is valid for $t \leq 1/2$:

$$\left| \frac{1}{\langle \widetilde{u}_i, v_i \rangle} - 1 \right| \leq 2\delta(1 + r\alpha\beta).$$

As a result,

$$\begin{aligned}
\|\bar{u}_i - \widetilde{u}_i\| = \|\widetilde{u}_i\| \left| \frac{1}{\langle \widetilde{u}_i, v_i \rangle} - 1 \right| \\
\leq 2\delta\sqrt{\alpha}(1 + r\alpha\beta) \\
\leq 4r\alpha^{3/2}\beta\delta.
\end{aligned}$$

Using again the norm equivalence bound and the triangle inequality,

$$\|\bar{U} - U\| \leq 5r^2\alpha^{3/2}\beta\delta, \tag{4.4}$$

which ends the preliminary part of the proof.

We now set accordingly $\bar{S} = \bar{U}\Sigma V^*$, and claim that S is now a truly diagonalizable matrix. Indeed, any $\bar{u}_i$ is an eigenvector of $\bar{S}$ with associated eigenvalue θ_i, and a basis of $\mathrm{im}\,(V)^{\perp}$ provides a family of eigenvectors of Σ with eigenvalue 0. We consequently set

$$\Pi = \begin{pmatrix} \bar{U} & Y \end{pmatrix},$$

where Y is an orthonormal basis of $\mathrm{im}\,(V)^{\perp}$; Π is the matrix of an eigenvector basis for S. Further, we have

$$\|\bar{S} - S\| \leq \|U - \bar{U}\|\|\Sigma\|\|V\| \leq 5r^2\alpha^2\beta\delta := \varepsilon'.$$

The above bound implies that the matrices A and $\bar{S}$ are still close:

$$\|A - \bar{S}\| \leq \|A - S\| + \|S - \bar{S}\| \leq \varepsilon + \varepsilon', \tag{4.5}$$

and we can apply the Bauer–Fike theorem to A and $\bar{S}$; the eigenvalues of A are contained in the union of the balls $B(\theta_i, \varepsilon'')$ and $B(0, \varepsilon'')$, where

$$\varepsilon'' = (\varepsilon + \varepsilon')\kappa(\Pi).$$

The computation of $\kappa(\Pi)$ being cumbersome, we defer the following lemma:

Lemma 1. *Let X be a $n \times r$ matrix with rank r, and X such that $X^*X' = I_r$. Let Y be a matrix for an orthonormal basis of $\mathrm{im}\,(X')^{\perp} = \ker((X')^*)$, and $P = (X, Y)$. Then, if $\|X\| \geq 1$ and $\|X'\| \geq 1$,*

$$\|\Pi\| \leq \sqrt{2}\|X\|, \quad \|\Pi^{-1}\| \leq \sqrt{2}(1 + \|X\|\|X'\|).$$

Applying this to $X = \bar{U}$ and $X' = V$ gives the bound

$$\kappa(\Pi) \leq 2\big(\|\bar{U}\| + \|\bar{U}\|^2 \|V\|\big),$$

and we use the triangle inequality to bound $\|\bar{U}\|$:

$$\|\bar{U}\| \leq \|U\| + \|\bar{U} - U\| \leq 6r^2\alpha^{3/2}\beta,$$

a very loose but sufficient bound, that entails

$$\kappa(\Pi) \leq 84r^2\alpha^{7/2}\beta.$$

The corresponding bound on ε'' reads

$$\varepsilon'' \leq 84r^2\alpha^{7/2}\beta(\varepsilon + 5r\alpha^2\beta\delta),$$

and we define σ to be the right-hand side of this inequality:

$$\sigma := 84r^2\alpha^{7/2}\beta(\varepsilon + 5r\alpha^2\beta\delta). \tag{4.6}$$

Going back to the Bauer–Fike application, the separation condition (4.2) implies that $B(0, \sigma)$ is disjoint from $B(\theta_i, \sigma)$ for $i \in [r]$ and we can apply the second part of the theorem: there are exactly r eigenvalues of A inside the region

$$\Omega = \bigcup_{i \in [r]} B(\theta_i, \sigma),$$

and all other eigenvalues of A have modulus less than σ. Further, again by the second part of Theorem 7, all connected components of Ω have the same number of eigenvalues of A and B. As a result, there exists a permutation π such that for all $i \in [r]$, we have

$$|\lambda_{\pi(i)} - \theta_i| \leq \sup_{\Omega' \subseteq \Omega} \text{diam}(\Omega') \leq 2r\sigma,$$

where the supremum is taken over all connected subsets of Ω.

We now move on to the eigenvector perturbation bound; let ξ be a normed eigenvector of A associated with the eigenvalue $\lambda_{\pi(i)}$. We write $\xi = \Pi x$ with Π the matrix defined before, and use (4.1):

$$\left\|\lambda_{\pi(i)}\Pi x - \sum_{j=1}^{r} \theta_j x_j \bar{u}_j\right\| = \left\|(A - \bar{S})x\right\| \leq \varepsilon + \varepsilon',$$

which we rewrite as

$$\left\|\Pi\left(\lambda_{\pi(i)}x - \sum_{j \in [r]} \theta_j x_j e_j\right)\right\| \leq \varepsilon + \varepsilon',$$

with $(e_1, \ldots, e_n)$ the usual orthonormal basis of $\mathbb{R}^n$. Using the inequality $\|v\| \leq \|P^{-1}\| \|Pv\|$ holding for any vector v,

$$\left\| \lambda_{\pi(i)} x - \sum_{j \in [r]} \theta_j x_j e_j \right\| \leq \|\Pi^{-1}\| (\varepsilon + \varepsilon').$$

We introduce the notation $\theta_{r+1} = \cdots = \theta_n = 0$; whenever the ball $B(\theta_i, \sigma)$ is disjoint from all other such balls, we have $|\lambda_{\pi(i)} - \theta_i| \leq \sigma$, and thus for $j \neq i$

$$|\lambda_{\pi(i)} - \theta_j| \geq |\theta_j - \theta_i| - |\lambda_{\pi(i)} - \theta_i| \geq \delta_i - \sigma,$$

so that

$$\|x - x_i e_i\| = \left\| \sum_{j \neq i} x_j e_j \right\| \leq \frac{1}{\delta_j - \sigma} \left\| \sum_{j \neq i} (\lambda_{\pi(i)} - \theta_j) x_j e_j \right\| \leq \frac{\|\Pi^{-1}\| (\varepsilon + \varepsilon')}{\delta_j - \sigma}.$$

We now apply Π inside the norm of the left-hand side, and use the fact that $\kappa(\Pi) \cdot (\varepsilon + \varepsilon') \leq \sigma$:

$$\|\xi - x_i \bar{u}_i\| \leq \frac{\sigma}{\delta_i - \sigma}.$$

Now, for any vectors $w, w' \in \mathbb{R}^n$, we have

$$\left\| \frac{w}{\|w\|} - \frac{w'}{\|w'\|} \right\| \leq \frac{2\|w - w'\|}{\|w\|}, \tag{4.7}$$

and all that remains is to write

$$\begin{aligned}
\left\| \xi - \frac{u_i}{\|u_i\|} \right\| &\leq \left\| \xi - \frac{\bar{u}_i}{\|\bar{u}_i\|} \right\| + \left\| \frac{u_i}{\|u_i\|} - \frac{\bar{u}_i}{\|\bar{u}_i\|} \right\| \\
&\leq \frac{2\sigma}{\delta_i - \sigma} + 2\|u_i - \bar{u}_i\| \\
&\leq \frac{3\sigma}{\delta_i - \sigma},
\end{aligned}$$

having used (4.7) twice and $2\|u_i - \bar{u}_i\| \leq \sigma$. This ends the proof. ■

As announced, we now prove the aforementioned Lemma 1 on the condition number of P:

Proof. Let $z \in \mathbb{R}^n$ be a unit vector, and write $z = \binom{x}{y}$ with x of size r and y of size $n - r$. Then, using that $\|Y\| = 1$,

$$\begin{aligned}
\|\Pi z\| = \|Xx + Yy\| &\leq \|X\| \cdot \|x\| + \|Y\| \cdot \|y\| \\
&\leq (1 \vee \|X\|)(\|x\| + \|y\|) \\
&\leq \sqrt{2}\|X\|,
\end{aligned}$$

which proves the first inequality. The second one relies on the following explicit formula for Π^{-1}:

$$\Pi^{-1} = \begin{pmatrix} (X')^* \\ -Y^*X(X')^* + Y^* \end{pmatrix}.$$

Indeed, using the relations $Y^*Y = I_{n-r}$ and $(X')^*Y = 0$:

$$\begin{aligned}
\begin{pmatrix} (X')^* \\ -Y^*X(X')^* + Y^* \end{pmatrix} P &= \begin{pmatrix} (X')^* \\ -Y^*X(X')^* + Y^* \end{pmatrix} (X \ Y) \\
&= \begin{pmatrix} (X')^*X & (X')^*Y \\ -Y^*X(X')^*X + Y^*X & -Y^*X(X')^*Y + Y^*Y \end{pmatrix} \\
&= \begin{pmatrix} I_r & 0 \\ -Y^*X + Y^*X & Y^*Y \end{pmatrix} \\
&= \begin{pmatrix} I_r & 0 \\ 0 & I_{n-r} \end{pmatrix} \\
&= I_n.
\end{aligned}$$

Furthermore, we have

$$\|-Y^*X(X')^* + Y^*\| \le \|Y\|\|I_n - X(X')^*\| \le 1 + \|X\|\|X'\|,$$

and the exact same argument as in the first inequality yields

$$\|P^{-1}\| \le \sqrt{2}\big(1 + \|X\|\|X'\|\big). \qquad \blacksquare$$

4.2. Matrix power perturbation and phase perturbation control

We aim in the following section to apply Theorem 8 to powers of the matrix B; however, such a process introduces uncertainty on the phase of the eigenvalues of B. The next theorem, adapted from [13] and [14], develops a method to control such uncertainty. As before, let $\Sigma = \text{diag}(\theta_1, \ldots, \theta_r)$ with

$$1 = |\theta_1| \ge \cdots \ge |\theta_r|,$$

and U, U', V, V' four $n \times r$ matrices. We set

$$S = U\Sigma^\ell V^*, \quad S' = U'\Sigma^{\ell'}(V')^*$$

for two integers ℓ, ℓ'.

Theorem 9. *Assume the following:*

(i)   *the integers ℓ, ℓ' are relatively prime,*

(ii) *the matrices U, U', V, V' are well-conditioned:*

- *they all are of rank r,*
- *for some $\alpha, \beta \geq 1$, for X in $\{U, V, U', V'\}$,*

$$\|X^*X\| \leq \alpha, \quad \|(X^*X)^{-1}\| \leq \beta,$$

- *for some small $\delta < 1$,*

$$\|U^*V - I_r\| \leq \delta, \quad \|(U')^*V' - I_r\| \leq \delta,$$

(iii) *there exists a small constant $\varepsilon > 0$ such that*

$$\|A^\ell - S\| \leq \varepsilon, \quad \|A^{\ell'} - S'\| \leq \varepsilon,$$

(iv) *if we let*

$$\sigma_0 := 84r^3\alpha^{7/2}\beta(\varepsilon + 5r\alpha^2\beta\delta),$$

then

$$\sigma_0 < \ell\,|\theta_r|^\ell, \quad \sigma_0 < \ell'\,|\theta_r|^{\ell'}. \tag{4.8}$$

Assume without loss of generality that ℓ is odd, and let

$$\sigma := \frac{\sigma_0}{\ell|\theta_r|^\ell}.$$

Then, the r largest eigenvalues of A are close to the θ_i in the following sense: there exists a permutation π of $[r]$ such that for $i \in [r]$,

$$|\lambda_{\pi(i)} - \theta_i| \leq 4\sigma,$$

and all other eigenvalues of A are less that $\sigma_0^{1/\ell}$. Additionally, if i is such that

$$B(\theta_i, \sigma) \cap \left(\bigcup_{j \neq i} B(\theta_j, \sigma)\right) = \emptyset, \tag{4.9}$$

then there exists a normed eigenvector ξ associated to $\lambda_{\pi(i)}$ such that

$$\left\|\xi - \frac{u_i}{\|u_i\|}\right\| \leq \frac{3\sigma}{\delta_i - \sigma},$$

with δ_i defined as in Theorem 8.

Proof. We apply Theorem 8 to A^ℓ, S and $A^{\ell'}$, S'; for any $i \in [r]$,

$$|\lambda^\ell_{\pi(i)} - \theta^\ell_i| \le \sigma_0, \quad |\lambda^{\ell'}_{\pi'(i)} - \theta^\ell_i| \le \sigma_0. \tag{4.10}$$

Examining the proof of Theorem 8, we notice that we can take $\pi = \pi'$ since taking the ℓ-th power does not change the ordering. We fix $i \in [r]$ and let $\lambda = \lambda_{\pi(i)} = |\lambda|e^{i\omega}$ and $\theta = \theta_i$ for now; then

$$\left| \frac{\lambda^\ell}{\theta^\ell} - 1 \right| \le \nu := \frac{\sigma_0}{|\theta|^\ell}.$$

The argument of $(\lambda/\theta)^\ell$ is thus between $-\xi$ and ξ, with

$$\xi = |2 \arcsin(\nu/2)| \le \pi/2\nu,$$

and the same holds for ℓ' (with ν' defined accordingly). Thus, there exists two integers p, p' and two numbers s, s' with absolute value less than $\pi/2\nu$ (resp. $\pi/2\nu'$), such that

$$\ell\omega = p\pi + s, \quad \ell'\omega = p'\pi + s'.$$

This implies

$$p\ell' - p'\ell = \frac{s'\ell - s\ell'}{\pi}.$$

The left-hand side of this inequality is an integer, and using condition (4.8) both terms in the right-hand side have a magnitude strictly lower than $1/2$, so both sides are 0. As ℓ and ℓ' are relatively prime, ℓ divides p and ℓ' divides p', so that

$$\omega = k\pi + \frac{s}{\ell}.$$

Whenever θ_i is positive, k is even and we can take $\omega = s/\ell$, and when k is odd we choose $\omega = \pi + s/\ell$.

We now come back to (4.10), and write

$$\lambda^\ell_i = \theta^\ell_i(1 + z)$$

with $|z| \le \nu$. Taking the modulus on both sides we find $|\lambda_i| = |\theta_i||1 + z|^{1/\ell}$ and we use the inequality $||1 + z|^{1/\ell} - 1| \le 2|z|/\ell$ (valid for $|z| \le 1/2$) to find

$$\left| |\lambda_i| - |\theta_i| \right| \le \frac{2\sigma_0}{\ell|\theta_i|^\ell}.$$

We can now prove the lemma: whether θ is positive or negative, a case analysis yields

$$\begin{aligned}
|\lambda_i - \theta_i| &\le \left| |\lambda_i| - |\theta_i| \right| + |\theta_i||e^{is/\ell} - 1| \\
&\le \frac{2\sigma_0}{\ell|\theta_i|^\ell} + |\theta_i|\frac{|s|}{\ell} \\
&\le \frac{4\sigma_0}{\ell|\theta_i|^\ell},
\end{aligned}$$

the desired bound. Now, assuming that ℓ is odd, we have by the mean value theorem

$$|\theta_i^\ell - \theta_j^\ell| \geq \ell\big(|\theta_i| \wedge |\theta_j|\big)^{\ell-1}|\theta_i - \theta_j| \geq \ell|\theta_r|^\ell|\theta_i - \theta_j|,$$

so that condition (4.9) implies the separation condition (4.3) applied to A^ℓ. We can then apply the same proof as in Theorem 8 and get

$$\left\|\xi - \frac{u_i}{\|u_i\|}\right\| \leq \frac{3\sigma_0}{\ell|\theta_r|^\ell\delta_i - \sigma_0},$$

which is equivalent to the theorem bound.   ∎

5. Proof of Theorem 2

We prove in this section the main result on the spectral properties of B. We shall use the same notations as in Theorem 2; since the statement of the theorem is invariant upon multiplying the entries of W by a common constant, we shall assume in the rest of the paper that $\mu_1 = 1$.

Our candidates for the singular vectors of B^ℓ are the vectors $(u_1, \ldots, u_{r_0})$ and $(v_1, \ldots, v_{r_0})$, where for $i \in [r_0]$,

$$u_i = \frac{B^\ell \chi_i}{\mu_i^\ell}, \quad v_i = \frac{(B^*)^\ell D_W \check{\chi}_i}{\mu_i^{\ell+1}}, \tag{5.1}$$

with associated eigenvalue μ_i^ℓ. We let U (resp. V) be the $n \times r$ matrix whose columns are the u_i (resp. v_i), and $D = \mathrm{diag}(\mu_1, \ldots, \mu_{r_0})$. The subspace spanned by the vectors $(v_1, \ldots, v_{r_0})$ will be denoted by H, and we let P_H and $P_{H^\perp}$ be the projections on H and its orthogonal $H^\perp$, respectively.

Finally, we will need an approximation of the Gram matrix of the vectors u (and v); we define for every $t \geq 0$ the *covariance* matrices $\Gamma_U^{(t)}$ and $\Gamma_V^{(t)}$ such that for $i, j \in [r_0]$,

$$\Gamma_{U,ij}^{(t)} = \sum_{s=0}^{t} \frac{\langle P\mathbf{1}, K^s\varphi^{i,j}\rangle}{(\mu_i\mu_j)^s}, \quad \Gamma_{V,ij}^{(t)} = \sum_{s=0}^{t} \frac{\langle K\mathbf{1}, K^s\varphi^{i,j}\rangle}{(\mu_i\mu_j)^{s+1}}, \tag{5.2}$$

where $\varphi^{i,j} = \varphi_i \circ \varphi_j$.

Remark. In the classical stochastic block model, we have $K = Q$ and the all-one vector $\mathbf{1}$ is an eigenvector of Q. This implies that the matrices $\Gamma_{U,ij}^{(t)}$ and $\Gamma_{V,ij}^{(t)}$ are diagonal, and thus the u_i (resp. v_i) are asymptotically orthogonal. This greatly simplifies the perturbation analysis of Theorem 8 for this special case.

5.1. Structure of the matrices U and V

Following from the subsequent local analysis of G, as well as a trace bound argument, we gather the following relations between matrices B^ℓ, D^ℓ and U.

Theorem 10. *Let r, d, b, τ, L be parameters as above, such that $a \leq n^{1/4}$, and (P, W) be any matrices in $\mathcal{C}(r, d, b, \tau, L)$. Let ℓ be any integer such that*

$$\ell \leq \frac{1 - \epsilon}{16} \frac{\log(n)}{\log(d)}, \tag{5.3}$$

for some $\epsilon > 0$. Then there exists an event with probability at least $1 - c/\log(n)$ and a parameter $N_0 \leq a^{12} L^6$ such that if $n \geq N_0$,

$$\|U^* U - \Gamma_U^{(\ell)}\| \leq C \times n^{-1/4}, \tag{5.4}$$

$$\|V^* V - \Gamma_V^{(\ell)}\| \leq C \times n^{-1/4}, \tag{5.5}$$

$$\|U^* V - I_{r_0}\|_\infty \leq C \times n^{-1/4}, \tag{5.6}$$

$$\|B^\ell U - U D^\ell\| \leq C'(\sqrt{\rho} \vee L)^\ell, \tag{5.7}$$

$$\|B^\ell P_{H^\perp}\| \leq C'(\sqrt{\rho} \vee L)^\ell, \tag{5.8}$$

where C and C' satisfy

$$C \leq c r d^4 b^2 L, \quad C' \leq c r^2 d^6 b^2 L^2 \log(n)^{20}.$$

Furthermore, on this same event, we have the following bound:

$$\|B^\ell\| \leq c \log(n) n^{1/4} L^\ell. \tag{5.9}$$

The proof of this theorem will occupy the next few pages of this article; we first show how it implies the statement of Theorem 2.

5.2. Proof of the perturbation bounds

The goal here is to apply Theorem 9 to B^ℓ, U and V: we choose ℓ equal to the upper bound in (5.3) (with arbitrary ϵ, say 0.01) and $\ell' = \ell + 1$, and let

$$S = U D^\ell V^*, \quad S' = U' D^{\ell'} V'^*,$$

where U', V' are defined identically to U and V replacing ℓ by ℓ'. We now check all the conditions of Theorem 9:

Condition (i). Since $\ell' = \ell + 1$, ℓ and ℓ' are relatively prime.

Condition (ii). We shall need a small lemma on the spectral properties of the covariance matrices, which will be proven in a subsequent section:

Lemma 2. *For all $t \geq 1$, the matrix $\Gamma_U^{(t)}$ (resp. $\Gamma_V^{(t)}$) is a positive definite matrix, with all its eigenvalues greater than 1 (resp. c_0^{-1}) and such that*

$$1 \leq \|\Gamma_U^{(t)}\| \leq \frac{r^2 d^3 L^2}{1 - \tau}, \quad c_0^{-1} \leq \|\Gamma_V^{(t)}\| \leq \frac{r^2 d^2 L^2}{1 - \tau}.$$

Then, the minimum eigenvalue of V^*V is at least $c_0^{-1} - Cn^{-1/4}$, which is more than $c_0^{-1}/2$ as soon as

$$n \geq c_1 r^4 a^4 d^{16} b^8 L^4,$$

and we can take $\beta = 2c_0$ whenever this holds. On the other hand,

$$\|V^*V\| \leq \frac{r d^2 L^2}{1 - \tau} + \frac{r b^2 d^4 L}{n^{1/4}} \leq \frac{2 r b^2 d^4 L^2}{1 - \tau}.$$

Performing the same computations on U^*U leads us to the choice

$$\alpha = \frac{2 r b^2 d^4 L^2}{1 - \tau}.$$

Finally, equation (5.6) allows us to take

$$\delta = Cn^{-1/4}.$$

Condition (iii). This condition requires some additional computations. Recall that $H = \text{im}(V)$; we have the formula

$$P_H = V(V^*V)^{-1}V^*.$$

Noticing that $SP_H = S$, we can bound $\|B^\ell - S\|$ as follows:

$$\begin{aligned}
\|B^\ell - S\| &\leq \|B^\ell P_H - SP_H\| + \|SP_{H\perp}\| + \|B^\ell P_{H\perp}\| \\
&\leq \|B^\ell P_H - S\| + \|B^\ell P_{H\perp}\| \\
&\leq \|B^\ell V(V^*V)^{-1} - UD\| \|V^*\| + \|B^\ell P_{H\perp}\|.
\end{aligned}$$

To apply (5.7), we let

$$U = P_H U + P_{H\perp} U = V(V^*V)^{-1} + \tilde{U} + P_{H\perp} U.$$

The second term is equal to $V(V^*V)^{-1}(V^*U - I_{r_0})$, and we can thus use (5.6):

$$\|\tilde{U}\| \leq \|V\| \|(V^*V)^{-1}\| \|V^*U - I_{r_0}\| \leq r \sqrt{\alpha} \beta \delta.$$

Going back to the above inequality, we find

$$\|B^\ell - S\| \le \|B^\ell U - UD^\ell\| + \|B^\ell\|\|\tilde{U}\| + \|B^\ell P_{H^\perp}\|\|U\| + \|B^\ell P_{H^\perp}\|,$$

and the bounds in Theorem 10 readily imply that all terms in the above inequality are bounded above by $\varepsilon := C''(\sqrt{\rho} \vee L)^\ell$, with

$$C'' \le \frac{c_3 r^3 a d^8 b^3 L^3 \log(n)^{20}}{1 - \tau}.$$

Condition (iv). Using all the bounds proven in the above computations, we find that

$$\sigma_0 \le C_0(\sqrt{\rho} \vee L)^\ell \quad \text{with } C_0 \le \frac{c_4 a^2 r^{11} d^{25} b^{13} L^{12} \log(n)^{20}}{(1 - \tau)^6}.$$

The bound we have to check is therefore

$$C_0(\sqrt{\rho} \vee L)^\ell \le \ell |\mu_{r_0}|^\ell \quad \Longleftrightarrow \quad C_0 \tau^\ell \le \ell,$$

which happens as soon as

$$\log(n) \ge \frac{20 \log(C_0) \log(d)}{\log(\tau^{-1})}.$$

The same proof holds for ℓ', with the same constants.

Having checked all assumptions of Theorem 9, we can now apply it to B^ℓ; this implies the existence of a permutation $\pi \in \mathfrak{S}_{r_0}$ (possibly depending on n) such that for $i \in [r_0]$,

$$|\lambda_i - \mu_{\pi(i)}| \le \sigma := C_0 \tau^\ell,$$

and all the other eigenvalues of B satisfy

$$|\lambda| \le C_0^{1/\ell}(\sqrt{\rho} \vee L).$$

Now, assume that for some $i \in [r_0]$, we have $\delta_i \ge 2\sigma$. Then, applying the last part of Theorem 9, there exists an eigenvector of B associated with λ_i such that

$$\left\| \xi - \frac{u_i}{\|u_i\|} \right\| \le \frac{3\sigma}{\delta_i - \sigma}.$$

We define in the following

$$\gamma_i = \langle P\mathbf{1}, (I_n - \mu_i^{-2} K)^{-1} \varphi^{i,i} \rangle.$$

If we rewrite the definition of $\Gamma_{U,ii}^{(t)}$ as

$$\Gamma_{U,ii}^{(t)} = \left\langle P\mathbf{1}, \sum_{s=0}^t (\mu_i)^{-2s} K^s \varphi^{i,j} \right\rangle,$$

the matrix sum converges as $t \to \infty$ since $\rho(K) < \mu_i^2$, and using Lemma 3 below, we have

$$|\Gamma_{U,ii}^{(\ell)} - \gamma_i| = \sum_{t=\ell+1}^{\infty} \mu_i^{-2t} \langle P\mathbf{1}, K^t \varphi^{i,j} \rangle \leq \sum_{t=\ell+1}^{\infty} rd^3 L^2 \rho^t \mu_i^{-2t} \leq \sigma,$$

and combined with (5.4) yields

$$\left| \|u_i\|^2 - \gamma_i \right| \leq 2\sigma.$$

On the other hand, we shall prove the following inequality in the following sections (equation (8.7)): for all $t \leq 2\ell$,

$$\left| \langle B^t \chi_i, \chi_i \rangle - \mu_i^t \langle P\mathbf{1}, \varphi^{i,i} \rangle \right| \leq \sigma \mu_i^t.$$

Setting $t = 0$ and $t = \ell$ in this inequality yields at the same time

$$\left| \|\chi_i\|^2 - \langle P\mathbf{1}, \varphi^{i,i} \rangle \right| \leq \sigma, \quad \left| \langle u_i, \chi_i \rangle - \langle P\mathbf{1}, \varphi^{i,i} \rangle \right| \leq \sigma.$$

We now have, using the Cauchy–Schwarz inequality,

$$\left| \langle \xi, \xi_i \rangle - \sqrt{\frac{\langle P\mathbf{1}, \varphi^{i,i} \rangle}{\gamma_i}} \right| \leq \left\| \xi - \frac{u_i}{\|u_i\|} \right\| + \left| \left\langle \frac{u_i}{\|u_i\|}, \frac{\chi_i}{\|\chi_i\|} \right\rangle - \sqrt{\frac{\langle P\mathbf{1}, \varphi^{i,i} \rangle}{\gamma_i}} \right|$$

$$\leq \frac{3\sigma}{\delta_i - \sigma} + c_5 \sigma \leq \frac{c_6 \sigma}{\delta_i - \sigma}.$$

Finally, notice that

$$\gamma_i = \langle P\mathbf{1}, \varphi^{i,i} \rangle + \left\langle P\mathbf{1}, \sum_{s=1}^{\infty} (\mu_i)^{-2s} K^s \varphi^{i,j} \right\rangle$$

$$\leq \langle P\mathbf{1}, \varphi^{i,i} \rangle + rd^2 L^2 \frac{\rho/\mu_i^2}{1 - \rho/\mu_i^2}.$$

Using that $rd^2 L^2 \geq 1$ and $\langle P\mathbf{1}, \varphi^{i,i} \rangle \geq 1$, we find

$$\frac{\langle P\mathbf{1}, \varphi^{i,i} \rangle}{\gamma_i} \geq 1 - rd^2 L^2 \frac{\rho}{\mu_i^2}.$$

6. Preliminary computations

We begin the proof of Theorem 10 with some elementary computations on the entries of K and $\Gamma^{(t)}$, which will be of use in the later parts of the proof. Most of the results from this section are adapted from [13], although sometimes improved and adapted to our setting.

Bounding ρ and L from below. We begin with a simple bound on $\rho = \rho(K)$; by the Courant–Fisher theorem, $\rho \geq \langle w, Kw \rangle$ for every unit vector w, and applying it to $w = 1/\sqrt{n}$ yields

$$\rho \geq \frac{\langle w, Kw \rangle}{n} = \frac{1}{n} \sum_{i,j \in [n]} P_{ij} \mathbb{E}[W_{ij}^2] = \frac{1}{d} \sum_{i,j \in [n]} P_{ij}^2 \mathbb{E}[W_{ij}]^2 = \frac{\|Q\|_F^2}{d},$$

where we used that $P_{ij} \leq d/n$ and the Jensen inequality. The Frobenius norm of Q is then greater than $\mu_1^2 = 1$, which in turns implies

$$\rho \geq \frac{1}{d}, \tag{6.1}$$

so that ρ is bounded away from zero. In order to prove a similar bound on L, we write for $x \in [n]$,

$$\varphi_1(x) = \sum_{y \in [n]} Q_{xy} \varphi_1(y) \leq \sqrt{\sum_y Q_{xy}^2} \leq \frac{dL}{\sqrt{n}}.$$

Squaring and summing those inequalities over x gives

$$1 = \|\varphi_1\|^2 \leq d^2 L^2,$$

so that as with ρ,

$$L \geq \frac{1}{d}. \tag{6.2}$$

A scalar product lemma. Our second step is an important lemma for the following proof, leveraging the entrywise bounds on W:

Lemma 3. *Let $\varphi, \varphi' \in \mathbb{R}^n$ be any unit vectors. Then, for any $t \geq 0$,*

$$\langle \mathbf{1}, K^t \varphi \circ \varphi' \rangle \leq r d^2 L^2 \rho^t.$$

Proof. We write the eigendecomposition of K as

$$K = \sum_{k=1}^{s} \nu_k \psi_k \psi_k^*,$$

with $\nu_1 = \rho$ the Perron–Frobenius eigenvalue of K and $s \leq r^2$ its rank. Then, for all $i \in [n]$,

$$\sum_{k=1}^{s} \nu_k^2 \psi_k(i)^2 = (K^2)_{ii} = \sum_{j \in [n]} K_{ij}^2$$

$$= \sum_{j \in [n]} P_{ij}^2 \mathbb{E}[W_{ij}^2]^2 \leq \sum_{j \in [n]} \left(\frac{d}{n}\right)^2 L^4 \leq \frac{d^2 L^4}{n}.$$

This is akin to a delocalization property on the eigenvectors of K.

We can now prove the above lemma:

$$\langle \mathbf{1}, K^t \varphi \circ \varphi' \rangle = \sum_{k=1}^{s} v_k^t \langle \mathbf{1}, \psi_k \rangle \langle \psi_k, \varphi \circ \varphi' \rangle$$

$$\leq \rho^{t-1} \sum_{k=1}^{s} \|\psi_k\| \|\mathbf{1}\| \cdot |v_k| |\langle \psi_k, \varphi \circ \varphi' \rangle|$$

$$\leq \rho^{t-1} \sqrt{n} \sum_{i \in [n]} |\varphi(i)| |\varphi'(i)| \sum_{k=1}^{s} |v_k| |\psi_k(i)|$$

$$\leq \rho^t d \sqrt{n} \sum_{i \in [n]} |\varphi(i)| |\varphi'(i)| \sqrt{s} \sqrt{\sum_{k=1}^{s} v_k^2 \psi_k(i)^2}$$

$$\leq \rho^t a \sqrt{n} \sqrt{s} \frac{dL^2}{\sqrt{n}} \sum_{i} |\varphi(i)| |\varphi'(i)| \leq r d^2 L^2 \rho^t,$$

where we extensively used the Cauchy–Schwarz inequality, as well as the bound $\rho^{-1} \leq d$ from (6.1). ∎

Entrywise bounds for K^t. For a more precise estimation of entrywise bounds, we define the scale-invariant *delocalization* parameter

$$\Psi = \frac{dL^2}{\rho}.$$

Using the same proof technique as in (6.2), as well as (6.1), we have

$$1 \leq \Psi \leq d^2 L^2$$

for any $i, j \in [n]$. Recall that, as shown in the proof of Lemma 3, for all $i \in [n]$,

$$(K^2)_{ii} \leq \frac{d^2 L^4}{n} = \frac{\Psi^2}{n} \rho^2.$$

Now, for $t \geq 0$ and $i, j \in [n]$,

$$(K^t)_{ij} = \sum_{k \in [s]} v_k^t \psi_k(i) \psi_k(j)$$

$$\leq \rho^{t-2} \sum_{k} v_k^2 |\psi_k(i)| |\psi_k(j)| \leq \rho^{t-2} \sqrt{(K^2)_{ii}(K^2)_{jj}},$$

where we again used the Cauchy–Schwarz inequality at the last line. This yields

$$(K^t)_{ij} \leq \frac{\Psi^2}{n} \rho^t \tag{6.3}$$

for any $t \geq 1$ and $i, j \in [n]$.

The covariance matrices. We now study the covariance matrices $\Gamma_U^{(t)}$ and $\Gamma_V^{(t)}$ defined in (5.2). Our aim is to prove the following lemma:

Lemma 4. *For all $t \geq 1$; the matrix $\Gamma_U^{(t)}$ (resp. $\Gamma_V^{(t)}$) is a positive definite matrix, with all its eigenvalues greater than 1 (resp. c^{-1}) and such that*

$$1 \leq \|\Gamma_U^{(t)}\| \leq \frac{r^2 d^3 L^2}{1 - \tau}, \quad c^{-1} \leq \|\Gamma_V^{(t)}\| \leq \frac{r^2 d^2 L^2}{1 - \tau}.$$

Proof. We first prove the bounds for $\Gamma_V^{(t)}$. Let $C^{(s)}$ be the $r_0 \times r_0$ matrix (where r_0 is defined as in Theorem 1) with

$$C_{ij}^{(s)} = \frac{\langle K\mathbf{1}, K^s \varphi^{i,j} \rangle}{(\mu_i \mu_j)^{s+1}}.$$

Then for every $w \in \mathbb{R}^{r_0}$, we have

$$w^* C^{(s)} w = \sum_{i,j \in [r_0]} \frac{w_i w_j}{(\mu_i \mu_j)^{s+1}} \sum_{x \in [n]} [K^{s+1} \mathbf{1}](x) \varphi_i(x) \varphi_j(x)$$

$$= \sum_{x \in [n]} [K^{s+1} \mathbf{1}](x) \left(\sum_{i \in [r_0]} \frac{w_i \varphi_i(x)}{\mu_i^{s+1/2}} \right)^2 \geq 0,$$

hence every matrix $C^{(s)}$ is positive semi-definite. Further, we have

$$C^{(0)} = D^{-1} \Phi^* \operatorname{diag}(K\mathbf{1}) \Phi D^{-1},$$

where Φ is the $n \times r$ matrix whose columns are the φ_i. Using $\mu_i \leq 1$ for any $i \in [r_0]$, the eigenvalues of $C^{(0)}$ are all greater than $\min_x [K\mathbf{1}](x) \geq c^{-1}$ by our initial assumptions. This settles the positive definite property, as well as the minimum eigenvalue of $\Gamma_V^{(t)}$.

Now, applying Lemma 3 to φ_i and φ_j, for all $i, j \in [r_0]$, one has

$$\Gamma_{V,ij}^{(t)} \leq \sum_{t=0}^{t} \frac{r d^2 L^2 \rho^{s+1}}{(\mu_i \mu_j)^{s+1}} \leq r d^2 L^2 \sum_{s=0}^{\infty} \left(\frac{\rho}{\mu_i \mu_j} \right)^s.$$

By definition of τ, the summand above is less than τ^s, whose sum converges since $\tau < 1$. As a result,

$$\|\Gamma_V^{(t)}\|_\infty \leq \frac{r d^2 L^2}{1 - \tau},$$

and the classic bound $\|\Gamma_V^{(t)}\| \leq r_0 \|\Gamma_V^{(t)}\|_\infty$ implies the upper bound.

The proof for $\Gamma_U^{(t)}$ is very similar; the upper bound simply ensues from the fact that $d_x \leq d$ for any $x \in [n]$. For the lower bound, if we let as above

$$C_{ij}'^{(s)} = \frac{\langle P\mathbf{1}, K^s \varphi^{i,j} \rangle}{(\mu_i \mu_j)^s},$$

then
$$C'^{(0)} = \Phi^* \operatorname{diag}(P\mathbf{1})\Phi,$$

and the minimum of $P\mathbf{1}$ is at least 1. This implies that the eigenvalues of $C'^{(0)}$ are larger than one, and we conclude as before. ∎

7. Local study of G

It is a well-known fact (see for example [14]) that when the mean degree is low enough ($d = n^{o(1)}$), the graph G is locally tree-like – that is, vertex neighborhoods behave almost like random trees. The goal of this section is to establish rigorously this result, as well as provide bounds on neighborhood sizes.

7.1. Setting and definitions

Labeled rooted graphs. A labeled rooted graph is a triplet $g_* = (g, o, \iota)$ consisting of a graph $g = (V, E)$, a root $o \in V$, and a mark function $\iota : V \to \mathbb{N}$ with finite support. We shall denote by $\mathcal{G}_*$ the set of labeled rooted graphs with $V = \mathbb{N}$, and will often write $g_* = (g, o)$ for an element of $\mathcal{G}_*$, dropping the mark function. Notions of subgraphs, induced subgraphs and distance extend naturally from regular graphs to this setting.

Labeling trees and graphs. We recall that G is the inhomogeneous random graph defined earlier. For each vertex $x \in V$, we can define the associated element of $\mathcal{G}_*$ as follows: the root is set to x, each vertex $y \in [n]$ is given a mark $\iota(y) = y$, and we let $\iota(z) = 0$ for all $z \in \mathbb{N} \setminus [n]$. The resulting triple (G, x, ι) is a random element of $\mathcal{G}_*$.

Now, let $o \in [n]$; we define the inhomogeneous random tree as follows: first, the root is given a mark $\iota(o) = o$. Then, for each vertex x already labeled, we draw the number of children of x according to $\operatorname{Poi}(d_{\iota(x)})$, where we recall that

$$d_{\iota(x)} = \sum_j P_{\iota(x),j} \leq d.$$

Each child y of x receives a label drawn independently at random from the distribution

$$\pi_{\iota(x)} = \left(\frac{P_{\iota(x),1}}{d_{\iota(x)}}, \ldots, \frac{P_{\iota(x),n}}{d_{\iota(x)}} \right), \tag{7.1}$$

which sums to 1 by definition. The resulting tree is a random element of $\mathcal{G}_*$, denoted by (T, o).

7.2. Growth properties of trees and graphs

A number of growth properties for neighborhoods in T and G are needed to ensure the successful couplings below. By definition of d, G (resp. (T, o)) is dominated by an Erdős–Rényi graph $\mathcal{G}(n, d/n)$ (resp. a Galton–Watson tree with offspring distribution $\mathrm{Poi}(d)$); we are thus able to directly lift properties from [14, Sections 8 and 9].

Lemma 5. *Let v be an arbitrary vertex in G; then, there exist absolute constants $c_0, c_1 > 0$ such that for every $s > 0$, we have*

$$\mathbb{P}\left(\forall t \geq 1, \ |\partial(G, v)_t| \leq sd^t\right) \geq 1 - c_0 e^{-c_1 s}. \tag{7.2}$$

The same result holds when replacing (G, v) with the tree (T, o) defined above.

Taking $s = c_1^{-1} \log(c_0 n^2)$ in the above inequality, one gets

$$\mathbb{P}\left(\forall t \geq 1, \ \forall v \in V, \ \left|\partial(G, v)_t\right| \leq c_3 \log(n) d^t\right) \geq 1 - \frac{1}{n}, \tag{7.3}$$

for any $n \geq 3$. Summing these inequalities for $1 \leq t \leq \ell$ yields a similar bound for the whole ball: with probability at least $1 - \frac{1}{n}$, we have

$$|(G, v)_t| \leq c_4 \log(n) d^t \tag{7.4}$$

for all $v \in V$ and $t \geq 1$. In particular, this implies the following useful bound: for any $v \in V$,

$$\deg(v) \leq c_4 d \log(n).$$

Another consequence of (7.2) is the following useful lemma:

Lemma 6. *For every $p \geq 2$, there is a constant c_p such that*

$$\mathbb{E}\left[\max_{v \in V} \sup_{t \geq 1} \left(\frac{|\partial(G, v)_t|}{d^t}\right)^p\right] \leq c_p \log(n)^p. \tag{7.5}$$

Similarly to the proof of (7.4), we have

$$\max_{v \in V} |(G, v)_t|^p \leq d^{tp} t^p \max_{x \in V} \sup_{s \leq t} \frac{|\partial(G, v)_t|^p}{d^{sp}},$$

which yields

$$\mathbb{E}\left[\max_{v \in V} |(G, v)_t|^p\right] \leq c_p t^p \log(n)^p d^{tp}. \tag{7.6}$$

An important note is that the above results apply to any collection of n random variables satisfying an inequality like (7.2); in particular, it also applies to an i.i.d. collection of inhomogeneous random trees of size n.

7.3. Local tree-like structure

We first check that the random graph G is tree-like. We say that a graph g is ℓ-*tangle-free* if there is at most one cycle in the ℓ-neighborhood of every vertex in the graph. As mentioned before, the random graph G is dominated by an Erdős–Rényi graph $\mathcal{G}(n, d/n)$; we can therefore lift the desired properties from [14].

Lemma 7. *Let $\ell \leq n$ be any integer parameter.*

   (i) *the random graph G is ℓ-tangle-free with probability at least $1 - ca^2 d^{4\ell}/n$.*

   (ii) *the probability that a given vertex v has a cycle in its ℓ-neighborhood is at most $cad^{2\ell}/n$.*

We shall assume in the following that the 2ℓ-tangle-free property happens with probability at least $1 - cn^{-\epsilon}$ for some $\epsilon > 0$, which happens whenever

$$\ell \leq \frac{1 - \epsilon}{10} \log_d(n) \leq c_3 \log(n). \tag{7.7}$$

We now gather all the result of the current section into one proposition, for ease of reading. The bound $\ell \leq c \log(n)$ assumed above is used to simplify the inequalities below.

Proposition 3. *Let G be an inhomogeneous random graph, and $(T_x, x)_{x\in[n]}$ a family of random trees as defined above. Let ℓ be small enough so that (7.7) holds. Then there exists an event $\mathcal{E}$ with probability at least $1 - (1/\log(n))$, under which:*

   (i) *the graph G is 2ℓ-tangle-free,*

   (ii) *for all $v \in G$, $t \leq 2\ell$, we have*

$$|(G, x)_t| \leq c \log(n) d^t, \tag{7.8}$$

   (iii) *for any $t \leq 2\ell$, the number of vertices in G whose t-neighborhood contains a cycle is at most $c \log(n)^2 d^{t+1}$.*

Furthermore, for any $t \leq 2\ell$ and $p \geq 1$, we have

$$\mathbb{E}\left[\max_{v \in V} |(G, v)_t|^p\right]^{1/p} \leq c \log(n)^2 d^t, \tag{7.9}$$

and the same holds for the family $(T_x, x)_{x\in[n]}$.

7.4. Coupling between rooted graphs and trees

We now turn onto the main argument of this proof: we bound the variation distance between the neighborhoods of (G, x) and (T, x) up to size ℓ.

First, recall some definitions: if $\mathbb{P}_1$, $\mathbb{P}_2$ are two probability measures on the space $(\Omega, \mathcal{F})$, their *total variation distance* is defined as

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{\mathcal{A} \in \mathcal{F}} |\mathbb{P}_1(\mathcal{A}) - \mathbb{P}_2(\mathcal{A})|.$$

The following two characterizations of the total variation distance shall be useful: first, whenever Ω is countable, we have

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \frac{1}{2}\|\mathbb{P}_1 - \mathbb{P}_2\|_1 = \frac{1}{2}\sum_{\omega \in \Omega} |\mathbb{P}_1(\omega) - \mathbb{P}_2(\omega)|. \tag{7.10}$$

Additionally,

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \min_{\mathbb{P} \in \pi(X_1, X_2)} \mathbb{P}(X_1 \neq X_2), \tag{7.11}$$

where $\pi(X_1, X_2)$ denotes the set of all *couplings* between $\mathbb{P}_1$ and $\mathbb{P}_2$, i.e. probability measures on $(\Omega^2, \mathcal{F} \otimes \mathcal{F})$ such that the marginal distributions are $\mathbb{P}_1$ and $\mathbb{P}_2$.

Denoting by $\mathcal{L}(X)$ the probability distribution of a variable X, the aim of this section is to prove the following:

Proposition 4. *Let $\ell \leq c_0 \log(n)$ for some constant $c_0 > 0$. Then, for every vertex $v \in V$,*

$$d_{\mathrm{TV}}\big(\mathcal{L}((G, v)_\ell), \mathcal{L}((T, v)_\ell)\big) \leq \frac{c \, \log(n)^2 d^{2\ell+2}}{n}. \tag{7.12}$$

7.4.1. A total variation distance lemma for sampling processes. For an integer n, denote by $\mathcal{S}(n)$ the set of all multisets with elements in $[n]$, and by $\mathcal{P}(n) \subset \mathcal{S}(n)$ the powerset of $[n]$. Let $p_1, \ldots, p_n \in [0, 1/2]$, with $\sum p_i = \lambda$ and $\sum p_i^2 = \alpha$, and consider the two probability laws on $\mathcal{S}(n)$:

- $\mathbb{P}_1$: each element i of $[n]$ is picked with probability p_i,
- $\mathbb{P}_2$: the size of the multiset S is drawn according to a $\mathrm{Poi}(\lambda)$ distribution, and each element of S has an i.i.d. label with distribution $(p_1/\lambda, \ldots, p_n/\lambda)$.

Note that $\mathbb{P}_1$ is actually supported on $\mathcal{P}(n)$.

Proposition 5. *Let $\mathbb{P}_1$, $\mathbb{P}_2$ be defined as above. Then*

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) \leq \alpha + \frac{e^{2\alpha} - 1}{2}.$$

Proof. Using characterization (7.10), we have

$$2 \, d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \sum_{S \in \mathcal{P}(n)} |\mathbb{P}_1(S) - \mathbb{P}_2(S)| + \mathbb{P}_2(S \notin \mathcal{P}(n)). \tag{7.13}$$

We shall treat those two terms separately. First, notice that for $S \in \mathcal{P}(n)$, we have

$$\mathbb{P}_1(S) = \prod_{i \in S} p_i \prod_{i \notin S} (1 - p_i) \tag{7.14}$$

$$\mathbb{P}_2(S) = \frac{e^{-\lambda} \lambda^{|S|}}{|S|!} \times |S|! \prod_{i \in S} \frac{p_i}{\lambda} = e^{-\lambda} \prod_{i \in S} p_i, \tag{7.15}$$

and thus by summing over all sets S,

$$\mathbb{P}_2(S \in \mathcal{P}(n)) = e^{-\lambda} \prod_{i=1}^{n} (1 + p_i).$$

Using the classical inequality $\log(1 + x) \geq x - x^2/2$, we can bound the second member of (7.13) as follows:

$$\begin{aligned}
\mathbb{P}_2(S \notin \mathcal{P}(n)) &= 1 - e^{-\lambda} \prod_{i=1}^{n} (1 + p_i) \\
&\leq 1 - e^{-\lambda} e^{\lambda - \alpha/2} \\
&\leq \alpha/2.
\end{aligned}$$

On the other hand, using again (7.14) and (7.15), the first term reduces to

$$\begin{aligned}
\sum_{S \in \mathcal{P}(n)} |\mathbb{P}_1(S) - \mathbb{P}_2(S)| &= \sum_{S \in \mathcal{P}(n)} \prod_{i \in S} p_i \left| \prod_{i \notin S} (1 - p_i) - e^{-\lambda} \right| \\
&\leq \sum_{S \in \mathcal{P}(n)} \prod_{i \in S} p_i \left(\left| e^{-\lambda} - \prod_{i=1}^{n} (1 - p_i) \right| + \left| \prod_{i \notin S} (1 - p_i) - \prod_{i=1}^{n} (1 - p_i) \right| \right).
\end{aligned}$$

Both absolute values above can be removed since the expressions inside are nonnegative; further, for $0 \leq p \leq 1/2$, we have $\log(1 - x) \geq -x - x^2$. Combining all those estimates, we find

$$\begin{aligned}
&\sum_{S \in \mathcal{P}(n)} |\mathbb{P}_1(S) - \mathbb{P}_2(S)| \\
&\leq e^{-\lambda}(1 - e^{-\alpha}) \sum_{S \in \mathcal{P}(n)} \prod_{i \in S} p_i + \prod_{i=1}^{n} (1 - p_i) \sum_{S \in \mathcal{P}(n)} \prod_{i \in S} p_i \left(\prod_{i \in S} \frac{1}{1 - p_i} - 1 \right) \\
&\leq \alpha e^{-\lambda} \prod_{i=1}^{n} (1 + p_i) + e^{-\lambda} \left(\prod_{i=1}^{n} \left(1 + \frac{p_i}{1 - p_i} \right) - \prod_{i=1}^{n} (1 + p_i) \right) \\
&\leq \alpha + e^{-\lambda} \exp\left(\sum_{i=1}^{n} \frac{p_i}{1 - p_i} \right) - e^{-\alpha/2},
\end{aligned}$$

where we again used the logarithm inequalities extensively. Finally, for $0 \leq p \leq 1/2$, we have $p/(1-p) \leq p + 2p^2$, which allows us to finish the computation:

$$\sum_{S \in \mathcal{P}(n)} |\mathbb{P}_1(S) - \mathbb{P}_2(S)| \leq \frac{3}{2}\alpha + e^{2\alpha} - 1. \tag{7.16}$$

Combining (7.16) with (7.13) easily implies the lemma. ∎

We introduce now a family of probability laws on $\mathcal{S}(n)$; for a subset $S \subseteq [n]$, let $\mathbb{P}_S$ be the measure corresponding to picking each element i of S with probability p_i.

The variation distance between those laws and $\mathbb{P}_1 = \mathbb{P}_{[n]}$ is then easier to bound:

Lemma 8. *For any $S \subseteq [n]$, we have*

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_S) \leq \sum_{i \notin S} p_i.$$

Proof. Consider the following coupling: we take a realization X of $\mathbb{P}_1$, and set $Y = X \cap S$. Then, $Y \sim \mathbb{P}_S$, and we find

$$\mathbb{P}(X \neq Y) = \mathbb{P}_1(X \cap S^c \neq \emptyset) \leq \mathbb{E}[|X \cap S^c|] = \sum_{i \notin S} p_i.$$

This ends the proof, since (7.11) ensures that $d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_S) \leq \mathbb{P}(X \neq Y)$. ∎

7.4.2. Proof of Proposition 4. Gathering all the previous results, we are now ready to prove Proposition 4.

Proof. Define the classical breadth-first exploration process on the neighborhood of a vertex v as follows : start with $A_0 = \{v\}$ and at stage $t \geq 0$, if A_t is not empty, take a vertex $v_t \in A_t$ at minimal distance from v, reveal its neighbors N_t in $V \setminus A_t$, and update $A_{t+1} = (A_t \cup N_t) \setminus \{v_t\}$. We denote by $(\mathcal{F}_t)_{t \geq 0}$ the filtration generated by the $(A_t)_{t \geq 0}$, and by $D_t = \bigcup_{s \leq t} A_s$ the set of vertices already visited at time t, and τ the first time at which all vertices in $(G, v)_\ell$ have been revealed.

We perform the same exploration process in parallel on (T, v), which corresponds to a breadth-first search of the tree. At step t, we denote by $\mathbb{P}_t$ the distribution of N_t given $\mathcal{F}_t$, and $\mathbb{Q}_t$ the distribution of the offspring of v_t in T (no conditioning is needed there).

Let E_ℓ denote the event that $(G, v)_\ell$ is a tree and contains no more than $c_1 \log(n) d^\ell$ vertices; from (7.4) and Lemma 7, we can choose c_1 such that E_ℓ has probability at least $1 - c_2 d^{2\ell+1}/n$ for some absolute constant c_2. By iteration, it suffices to show that if E_ℓ holds, there exists a constant $c_3 > 0$ such that

$$d_{\mathrm{TV}}(\mathbb{P}_t, \mathbb{Q}_t) \leq \frac{c_3 \log(n) d^{\ell+2}}{n} \quad \text{for all } t \leq \tau. \tag{7.17}$$

Given $\mathcal{F}_t$, the probability measure $\mathbb{P}_t$ is as follows: each element i of $V \setminus A_t$ is selected with probability $p_i = P_{v_t i}$. Let $\mathbb{P}'_t$ denote the same probability measure, but where the selection is made over all of V. Using Lemma 8, we first find that

$$d_{\mathrm{TV}}(\mathbb{P}_t, \mathbb{P}'_t) \leq \sum_{i \in A_t} P_{v_t i} \leq c_1 \log(n) d^\ell \cdot \frac{d}{n}.$$

On the other hand, Proposition 5 yields

$$d_{\mathrm{TV}}(\mathbb{P}'_t, \mathbb{Q}_t) \leq c_4 \sum_{i=1}^{n} P_{v_t i}^2 \leq c_5 \frac{d^2}{n}.$$

Equation (7.17) then results from a straightforward application of the triangle inequality. ∎

8. Near eigenvectors of G

8.1. Functionals on (T, o)

8.1.1. Vertex functionals on trees. Similarly to [14], quantities of interest in the study of B will be tied to functionals on the random inhomogeneous tree defined above. Define a functional $f_{\varphi,t}$ on the set of labeled rooted trees $\mathcal{T}_* \subset \mathcal{G}_*$ by

$$f_{\varphi,t}(T, o) = \sum_{x_t \in \partial(T,o)_t} W_{\iota(o),\iota(x_1)} \ldots W_{\iota(x_{t-1}),\iota(x_t)} \varphi(\iota(x_t)),$$

where $(o, x_1, \ldots, x_t)$ is the unique path of length t between o and x_t. Then the following proposition holds:

Proposition 6. *Let $t \geq 0$ be an integer. For any $i, j \in [r]$, the following identities are true:*

$$\mathbb{E}\big[f_{\varphi_i,t}(T, x)\big] = \mu_i^t \varphi_i(x), \tag{8.1}$$

$$\mathbb{E}\big[f_{\varphi_i,t}(T, x) f_{\varphi_j,t}(T, x)\big] = (\mu_i \mu_j)^t \sum_{s=0}^{t} \frac{[K^s \varphi^{i,j}](x)}{(\mu_i \mu_j)^s}, \tag{8.2}$$

$$\mathbb{E}\big[(f_{\varphi_i,t+1}(T, x) - \mu_i f_{\varphi_i,t}(T, x))^2\big] = [K^{t+1} \varphi^{i,i}](x), \tag{8.3}$$

where we recall that $\varphi^{i,j} = \varphi_i \odot \varphi_j$.

8.1.2. Adapting functionals to non-backtracking paths. The matrix B considered here acts on (directed) edges, whereas the functionals considered so far are defined on vertices. Consequently, we define the following transformation: for a function $f : \mathcal{G}_* \to \mathbb{R}$, and a random vector $w \in \mathbb{R}^V$ with expected value $\bar{w}$, let

$$\vec{\partial}_w f(g, o) = \sum_{e : e_2 = o} w_{e_1} f(g_e, o),$$

where g_e denotes the graph g with the edge e_1, e_2 removed.

The expectations from Proposition 6 are then adapted as follows:

Proposition 7. *Let $t \geq 0$ be an integer. For any $i, j \in [r]$, and $\phi \in \ker(P)$, the following identities are true:*

$$\mathbb{E}\big[\vec{\partial}_w f_{\varphi_i, t}(T_x, x)\big] = [P\bar{w}](x) \cdot \mathbb{E}\big[f_{\varphi_i, t}(T_x, x)\big], \tag{8.4}$$

$$\mathbb{E}\big[\vec{\partial}_w (f_{\varphi_i, t} \cdot f_{\varphi_j, t})(T_x, x)\big] = [P\bar{w}](x) \cdot \mathbb{E}\big[f_{\varphi_i, t}(T_x, x) f_{\varphi_j, t}(T_x, x)\big], \tag{8.5}$$

$$\mathbb{E}\big[\vec{\partial}_w [(f_{\varphi_i, t+1} - \mu_i f_{\varphi_i, t})^2](T_x, x)\big]$$
$$= [P\bar{w}](x) \cdot \mathbb{E}\big[(f_{\varphi_i, t+1}(T_x, x) - \mu_i f_{\varphi_i, t}(T_x, x))^2\big]. \tag{8.6}$$

The proof for those results makes use of properties specific to moments of Poisson random variables; as with the preceding results, it is deferred to a later section.

8.2. Spatial averaging of graph functionals

In this section, we leverage the coupling obtained above to provide bounds on quantities of the form $\frac{1}{n} \sum_{x \in V} f(G, x)$, for local functions f. The tools and results used in this section are essentially identical to those in [14], with a few improvements and clarifications added when necessary.

We begin with a result that encodes the fact that the t-neighborhoods in G are approximately independent. We say that a function f from $\mathcal{G}_*$ to $\mathbb{R}$ is t-local if $f(g, o)$ is only function of $(g, o)_t$.

Proposition 8. *Let $t \leq c_0 \log(n)$ for some constant $c_0 > 0$. Let $f, \psi : \mathcal{G}_* \to \mathbb{R}$ be two t-local functions such that $|f(g, o)| \leq \psi(g, o)$ for all $(g, o) \in \mathcal{G}_*$ and ψ is non decreasing by the addition of edges. Then*

$$\mathrm{Var}\Big(\sum_{o \in V} f(G, o)\Big) \leq c \log(n)^4 n d^{2t} \cdot \sqrt{\mathbb{E}\big[\max_{o \in V} \psi(G, o)^4\big]}.$$

Proof. For $x \in V$, denote by E_x the set $\{\{u, x\} \in E \mid u \leq x\}$; the vector $(E_1, \ldots, E_n)$ is an independent vector, and we have

$$Y := \sum_{v \in V} f(G, v) = F(E_1, \ldots, E_n)$$

for some measurable function F.

Define now G_x the graph with vertex set V and edge set $\bigcup_{y \neq x} E_y$, and set

$$Y_x = \sum_{v \in V} f(G_x, v).$$

The random variable Y_x is $\bigcup_{y \neq x} E_y$-measurable, so the Efron–Stein inequality applies:

$$\mathrm{Var}(Y) \leq \sum_{x \in [n]} \mathbb{E}\big[(Y - Y_x)^2\big].$$

For a given $x \in V$, the difference $f(G, o) - f(G_x, o)$ is always zero except if $x \in (G, o)_t$, due to the locality property; consequently,

$$
\begin{aligned}
|Y - Y_x| &\leq \sum_{o \in V} |f(G, o) - f(G_x, o)| \\
&\leq \sum_{o \in (G, x)_t} \psi(G, o) + \psi(G_x, o) \\
&\leq 2 \max_{x \in [n]} |(G, x)_t| \cdot \max_{o \in V} \psi(G, o),
\end{aligned}
$$

where we used the non-decreasing property of ψ in the last line. By the Cauchy–Schwarz inequality and equation (7.6), we can write

$$
\begin{aligned}
\mathbb{E}\big[(Y - Y_x)^2\big] &\leq 4 \sqrt{\mathbb{E}\big[|\max_{x \in [n]} (G, x)_t|^4\big]} \cdot \sqrt{\mathbb{E}\big[\max_{o \in V} \psi(G, o)^4\big]} \\
&\leq c_1 t^2 \log(n)^2 d^{2t} \cdot \sqrt{\mathbb{E}\big[\max_{o \in V} \psi(G, o)^4\big]}.
\end{aligned}
$$

Using that $t \leq c_0 \log(n)$, and the linearity of expectation, yields the desired bound. ∎

We now use our previous coupling results to provide a concentration bound between a functional on graphs and its expectation on trees:

Proposition 9. *Let* $t \in \mathbb{N}$ *and* $f, \psi \colon \mathcal{G}_* \to \mathbb{R}$ *be as in the previous proposition. Then, with probability at least* $1 - (1/(r^2 \log(n)^2))$, *the following inequality holds:*

$$\left| \sum_{v \in V} f(G, v) - \mathbb{E}\left[\sum_{x \in [n]} f(T_x, x) \right] \right| \leq c\, r \log(n)^3 d^{t+1} \sqrt{n}\, \|\psi\|_\star,$$

where $\|\psi\|_\star$ *is defined as*

$$\|\psi\|_\star = \big(\mathbb{E}\big[\max_{v \in V} \psi(G, v)^4\big]\big)^{1/4} \vee \big(\max_{x \in [n]} \mathbb{E}\big[\psi(T_x, x)^2\big]\big)^{1/2}.$$

Proof. Using the Chebyshev inequality and the variance bound from the preceding proposition, with probability at least $1 - (1/(r^2 \log (n)^2))$, we have

$$\left| \sum_{v \in V} f(G, v) - \mathbb{E}\left[\sum_{v \in V} f(G, v) \right] \right| \le c_1 \, r \log (n)^3 d^t \sqrt{n} \|\psi\|_\star.$$

It then remains to bound the difference between the expectation term and its counterpart on trees. For $x \in V$, let $\mathcal{E}_x$ denote the event that the coupling between $(G, x)_t$ and $(T_x, x)_t$ fails; by the locality property, $f(G, x) = f(T_x, x)$ on $\mathcal{E}_x$. Therefore, using the Cauchy–Schwarz inequality,

$$\left| \sum_{x \in [n]} \mathbb{E}\big[f(G, x) - f(T_x, x) \big] \right| \le \sum_{x \in [n]} \mathbb{E}\big[|f(G, x)| 1_{\mathcal{E}_x} + |f(T_x, x)| 1_{\mathcal{E}_x} \big]$$

$$\le \sum_{x \in [n]} \sqrt{\mathbb{P}(*)\mathcal{E}_x} \Big(\sqrt{\mathbb{E}\big[\psi(G, x)^2 \big]} + \sqrt{\mathbb{E}\big[\psi(T_x, x)^2 \big]} \Big)$$

$$\le \sqrt{\frac{c_2 \log (n)^2 d^{2t+2}}{n}} \cdot \sum_{x \in [n]} \Big(\mathbb{E}\big[\psi(G, x)^4 \big]^{1/4} + \sqrt{\mathbb{E}\big[\psi(T_x, x)^2 \big]} \Big)$$

$$\le c_3 \log(n) a d^{t+1} \sqrt{n} \, \|\psi\|_\star.$$

It is then straightforward to check that both obtained bounds are less than the right-hand side in the proposition, upon adjusting c. ∎

8.3. Structure of near eigenvectors

In the following, the aim is to obtain bounds on the norms and scalar product of the near eigenvectors u_i and v_i defined in (5.1). The main result of this section is as follows:

Proposition 10. *Let ℓ be small enough so that* (7.7) *holds. On an event with probability* $1 - c_1 / \log(n)$, *the following inequalities hold for all $i, j \in [r]$, $t \le 2\ell$ and some absolute constant $c > 0$:*

$$\left| \langle B^t \chi_i, \chi_j \rangle - \mu_i^t \langle \varphi_i, D_P \varphi_j \rangle \right| \le \frac{c \, r b^2 d^2 \log (n)^6 d^{2t} L^t}{\sqrt{n}}, \tag{8.7}$$

$$\left| \langle B^t \chi_i, D_W \check{\chi}_j \rangle - \mu_i^{t+1} \delta_{ij} \right| \le \frac{c \, r b^2 d^3 L \log (n)^6 d^{2t} L^t}{\sqrt{n}}, \tag{8.8}$$

$$\left| \langle B^t \chi_i, B^t \chi_j \rangle - \mu_i^t \mu_j^t \Gamma_{U,ij}^{(t)} \right| \le \frac{c \, r b^2 d^2 \log (n)^7 d^{3t} L^{2t}}{\sqrt{n}}, \tag{8.9}$$

$$\left| \langle (B^*)^t D_W \check{\chi}_i, (B^*)^t D_W \check{\chi}_j \rangle - \mu_i^{t+1} \mu_j^{t+1} \Gamma_{ij}^{(t+1)} \right|$$
$$\leq \frac{c \, r b^2 d^2 L^2 \log(n)^6 d^{3t} L^{2t}}{\sqrt{n}}, \tag{8.10}$$

$$\left\| B^{t+1} \chi_i - \mu_i B^t \chi_i \right\|^2 \leq r d^3 L^2 \rho^{t+1} + \frac{c r b^2 d^3 \log(n)^7 d^{3t} L^{2t}}{\sqrt{n}}. \tag{8.11}$$

Proof. The proof of those inequalities relies on careful applications of Proposition 9 to previously considered functionals. We aim to prove that each of those inequalities hold with probability $1 - c_2 / r \log(n)$; we fix in the following an integer $t \leq 2\ell$ and $i, j \in [r]$. Let $\mathcal{V}_t$ be the set of vertices such that $(G, v)_t$ is not a tree; we place ourselves in the event described in Proposition 3 and as a consequence

$$\mathcal{V}_t \leq c_3 \log(n)^2 d^{t+1}.$$

We first prove (8.7); let

$$f(g, o) = \mathbf{1}_{(g,o)_t \text{ has no cycles}} \varphi_j(o) \vec{\partial}_1 f_{\varphi_i, t}(g, o).$$

The function f is clearly t-local, and

$$|f(g, o)| \leq \|\varphi_i\|_\infty \|\varphi_j\|_\infty \deg(o) |\partial(g, o)_t| L^t$$
$$\leq \frac{b^2}{n} \deg(o) |(g, o)_t| L^t := \psi(g, o).$$

The function ψ thus defined is non-decreasing by the addition of edges. When $v \notin \mathcal{V}_t$, we notice that

$$f(G, v) = \varphi_j(v) \cdot [T^* B^t \chi_i](v),$$

hence,

$$\left| \langle B^t \chi_i, \chi_j \rangle - \sum_{v \in V} f(G, v) \right| = \left| \sum_{v \in \mathcal{V}_t} \varphi_i(v) T^* B^t \chi_j \right| \leq 2 |\mathcal{V}_t| \max_v \psi(G, v),$$

since by the tangle-free property there are at most two paths from v to any vertex in $(G, v)_t$. Furthermore, using the results in Section 7.2, we find that with probability at least $1 - 1/n$,

$$\max_v \psi(G, v) \leq \frac{c_4 b^2 \log(n)^2 d^{t+1} L^t}{n} \psi\|_\star \leq \frac{c_4 b^2 \log(n)^3 d^{t+1} L^t}{n}.$$

Finally, a direct computation shows that

$$\sum_{x \in [x]} \mathbb{E}\big[f(T_x, x) \big] = \sum_{x \in [n]} \varphi_j(x) \cdot d_x \mu_i^t \varphi_i(x) = \mu_i^t \langle \varphi_j, D_P \varphi_i \rangle.$$

Applying Proposition 9 to f and ψ, and using the triangle inequality:

$$\left|\langle B^t \chi_i, \chi_j \rangle - \mu_i^t \langle \varphi_j, D_P \varphi_i \rangle\right| \leq \frac{c_5 \, b^2 \log (n)^4 d^{2t+2} L^t}{n} + \frac{c_6 \, r b^2 \log (n)^6 d^{2t+2} L^t}{\sqrt{n}}$$

$$\leq \frac{c_7 \, r b^2 d^2 \log (n)^6 d^{2t} L^t}{\sqrt{n}}.$$

The proof of the other inequalities is very similar, applying Proposition 9 to other functionals from Section 8.1. To avoid clutter, it is deferred to the appendix. ∎

9. Proof of Theorem 10

Having shown Proposition 10, all that remains is simply to gather the preceding bounds, and simplify them to get an easy-to-read summary. Bounds (5.4)–(5.6), as well as (5.9), being straightforward computations, they are deferred to the appendix.

9.1. A telescopic trick: proof of (5.7)

Notice that for a $r_0 \times r_0$ matrix M, we have

$$\|M\| \leq r_0 \max_i \|M_i\|, \tag{9.1}$$

where M_i are the columns (or lines) of M. To apply this inequality, we write

$$\|B^\ell u_i - \mu_i^\ell u_i\| \leq \sum_{t=0}^{\ell-1} \mu_i^{\ell-t-1} \|B^{t+1} u_i - \mu_i B^t u_i\|, \tag{9.2}$$

and (8.11) yields

$$\|B^{t+1} u_i - \mu_i B^t u_i\|^2 \leq \mu_i^{-2\ell} \|B^{t+\ell+1} \chi_i - \mu_i B^{t+\ell} \chi_i\|^2$$

$$\leq \mu_i^{-2\ell}\left(r d^3 L^2 \rho^{t+\ell+1} + \frac{c r b^2 d^3 \log (n)^7 d^{3(t+\ell)} L^{2(t+\ell)}}{\sqrt{n}}\right).$$

Since $i \leq r_0$, the bounds $\mu_i^2 \geq \rho \geq 1/d$ apply, so that

$$\|B^{t+1} u_i - \mu_i B^t u_i\|^2 \leq r d^3 L^2 \rho^{t+\ell+1} \mu_i^{-2\ell}$$

$$+ \frac{c r b^2 d^3 \log (n)^7 d^{3t+5\ell} L^{2(t+\ell)}}{\sqrt{n}}. \tag{9.3}$$

We now use the (very crude) inequality $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ inside (9.3):

$$\|B^\ell u_i - \mu_i^\ell u_i\|$$

$$\leq \sum_{t=0}^{\ell-1} \left[\mu_i^{\ell-t-1} \sqrt{r} d^{3/2} L \rho^{(t+\ell+1)/2} \mu_i^{-\ell} + \frac{c_1 \, bd^{3/2} \log{(n)}^{7/2} d^{(3t+5\ell)/2} L^{t+\ell}}{n^{1/4}} \right]$$

$$\leq \sqrt{r} d^{3/2} L \rho^{\ell/2} \sum_{t=0}^{\ell-1} \left(\frac{\sqrt{\rho}}{\mu_i} \right)^{t+1} + c_2 \, bd^2 \log{(n)}^{9/2} \frac{(Ld^4)^\ell}{n^{1/4}} L^\ell.$$

The terms in the sum are all less than 1 since $i \leq r_0$, and $\ell < c_3 \log(n)$ implies

$$\|B^\ell u_i - \mu_i^\ell u_i\| \leq c_3 \sqrt{r} d^{3/2} L \log(n) \rho^{\ell/2} + c_2 bd^2 \log{(n)}^{9/2} \frac{(aLd^3)^\ell}{n^{1/4}} L^\ell.$$

The bound $(Ld^4)^\ell \leq n^{1/4}$ holds by definition of ℓ, and (5.7) ensues via (9.1).

9.2. Bounding $\|B^\ell P_{H^\perp}\|$

Having established the candidates and error bounds for the upper eigenvalues of B^ℓ, it remains to bound the remaining eigenvalues (also called the *bulk*) of the matrix. This is done using a method first employed in [28], and leveraged again in a similar setting in [13, 14]. Our approach will be based on the latter two, adapting the non-backtracking method to the weighted case.

Our first preliminary step is the following lemma:

Lemma 9. *On an event with probability at least $1 - 1/\log(n)$, for any $t \leq \ell$, any unit vector $w \in H^\perp$ and $i \in [r_0]$, one has*

$$\left| \langle (B^*)^t D_W \check{\chi}_i, w \rangle \right| \leq \sqrt{r} d^{3/2} L^2 \rho^{t/2} + \frac{c_4 \, bd^{3/2} \log{(n)}^{9/2} d^{2\ell} L^\ell}{n^{1/4}}.$$

Proving this bound is done through the same telescopic sum trick as above, and is done in the appendix.

9.2.1. Tangle-free decomposition of B^ℓ.
We adapt here the decomposition first used in [14] to our setting. Through the remainder of this section, we shall consider B as an operator on $\vec{E}(V)$ instead of $\vec{E}$, setting $B_{ef} = 0$ whenever $e \notin \vec{E}$ or $f \notin \vec{E}$. This yields a matrix with B as a principal submatrix and zeros everywhere else, thus the non-zero spectrum stays identical.

For $e, f \in \vec{E}(V)$, and $t \geq 0$, we define Γ_{ef}^k the set of non-backtracking paths of length k from e to f; further, for an edge e we define X_e the indicator variable of $e \in \vec{E}$, and $A_e = X_e W_e$, so that A is the (weighted) adjacency matrix of G.

We then have that

$$(B^k)_{ef} = \sum_{\gamma \in \Gamma_{ef}^{k+1}} X_e \prod_{s=1}^{k} A_{\gamma_s \gamma_{s+1}}.$$

Define F_{ef}^k the set of ℓ-tangle-free paths (i.e. the set of paths γ such that the subgraph induced by γ is tangle-free). Then, whenever the graph G is tangle-free, for all $k \le \ell$ the matrix B^k is equal to $B^{(k)}$, with

$$(B^{(k)})_{ef} = \sum_{\gamma \in F_{ef}^{k+1}} X_e \prod_{s=1}^{k} A_{\gamma_s \gamma_{s+1}}.$$

Define now the "centered" versions of the weighted and unweighted adjacency matrices $\underline{A}$ and $\underline{X}$ by

$$\underline{A}_{ij} = A_{ij} - Q_{ij}, \quad \underline{X}_{ij} = X_{ij} - P_{ij}$$

for every $i \ne j$, and its centered non-backtracking counterpart as

$$(\Delta^{(k)})_{ef} = \sum_{\gamma \in F_{ef}^{k+1}} \underline{X}_{ij} \prod_{s=1}^{k} \underline{A}_{\gamma_s \gamma_{s+1}},$$

with the convention that the product over an empty set is equal to 1.

Recall that for any two sets of real numbers $(x_i), (y_i)$, we have the following:

$$\prod_{s=0}^{\ell} x_s = \prod_{s=0}^{\ell} y_s + \sum_{t=0}^{\ell} \prod_{s=0}^{t-1} y_s (x_t - y_t) \prod_{s=t+1}^{\ell} x_s.$$

Applying this formula to the above definitions, and separating the case $t = 0$ in the sum yields

$$B_{ef}^{(\ell)} = \Delta_{ef}^{(\ell)} + \sum_{\gamma \in F_{ef}^{\ell+1}} Q_e \prod_{s=1}^{\ell} \underline{A}_{\gamma_s \gamma_{s+1}}$$

$$+ \sum_{t=1}^{\ell} \sum_{\gamma \in F_{ef}^{\ell+1}} \underline{X}_e \prod_{s=1}^{t-1} \underline{A}_{\gamma_s \gamma_{s+1}} Q_{\gamma_t \gamma_{t+1}} \prod_{s=t+1}^{\ell} A_{\gamma_s \gamma_{s+1}}. \tag{9.4}$$

Define now $F_{t,ef}^{\ell+1} \subset \Gamma_{ef}^{\ell+1}$ the set of non-backtracking *tangled* paths γ such that $(\gamma_0, \dots, \gamma_t) \in F_{eg}^t$, $(\gamma_{t+1}, \dots, \gamma_{\ell+1}) \in F_{g'f}^{\ell-t}$ for some edges $g, g' \in \vec{E}(V)$. As an edge case, $F_{0,ef}^{\ell+1}$ is the set of tangled paths γ such that $(\gamma_0, \gamma_1) = e_1$ and $(\gamma_1, \dots, \gamma_{\ell+1}) \in F_{g'f}^{\ell}$ for some $g' \in \vec{E}(V)$ (note that necessarily $e_2 = g'_1$), and similarly for $F_{\ell,ef}$.

Finally, we introduce the two matrices M and $M^{(2)}$ as

$$M_{ef} = \mathbf{1}\{e \to f\}Q_e, \quad M_{ef}^{(2)} = \mathbf{1}(e \xrightarrow{2} f)Q_{e_2 f_1}$$

for $e, f \in \vec{E}(V)$, where $e \xrightarrow{2} f$ means that there exists a non-backtracking path of length two between e and f. Then, equation (9.4) can be rewritten as

$$B^{(\ell)} = \Delta^{(\ell)} + MD_W B^{(\ell-1)}$$
$$+ \sum_{t=1}^{\ell-1} \Delta^{(t-1)} M^{(2)} D_W B^{(\ell-t-1)} + \Delta^{(\ell-1)} M - \sum_{t=0}^{\ell} R_t^{(\ell)}, \qquad (9.5)$$

where

$$(R_t^{(\ell)})_{ef} = \sum_{\gamma \in F_{t,ef}^{\ell+1}} \underline{X}_e \prod_{s=1}^{t-1} \underline{A}_{\gamma_s \gamma_{s+1}} Q_{\gamma_t \gamma_{t+1}} \prod_{s=t+1}^{\ell} A_{\gamma_s \gamma_{s+1}},$$

$$(R_0^{(\ell)})_{ef} = \sum_{\gamma \in F_{t,ef}^{\ell+1}} Q_e \prod_{s=1}^{\ell} A_{\gamma_s \gamma_{s+1}}.$$

Note that $M^{(2)}$ is pretty close to a modified version of Q; more specifically, we make the decomposition

$$M^{(2)} = TQT^* + \tilde{M} = \sum_{k=1}^{r} \mu_k \chi_k \check{\chi}_k^* + \tilde{M}.$$

Then, the following decomposition holds:

$$B^{(\ell)} = \Delta^{(\ell)} + MD_W B^{(\ell-1)} + \sum_{t=1}^{\ell-1} \sum_{k=1}^{r} \mu_k \Delta^{(t-1)} \chi_k \check{\chi}_k^* D_W B^{(\ell-t-1)}$$
$$+ \sum_{t=1}^{\ell-1} \Delta^{(t-1)} \tilde{M} B^{(\ell-t-1)} + \Delta^{(\ell-1)} M - \sum_{t=0}^{\ell} R_t^{(\ell)}.$$

Noticing that $\|M\| \leq d$ and $\|\chi_k\| \leq d \log(n)$, the following lemma ensues:

Lemma 10. *On an event with probability at least $1 - 1/\log(n)$, the following inequality holds for any normed vector $x \in \mathbb{R}^{\vec{E}(V)}$:*

$$\|B^{\ell}x\| \leq \|\Delta^{(\ell)}\| + L\|MB^{\ell-1}\| + d\log(n) \sum_{t=1}^{\ell-1} \|\Delta^{(t-1)}\| \sum_{k=1}^{r} |\langle D_W \check{\chi}_k, B^{\ell-t-1}x \rangle|$$
$$+ \sum_{t=1}^{\ell-1} \|\Delta^{(t-1)} \tilde{M} B^{\ell-t-1}\| + d\|\Delta^{(\ell-1)}\| - \sum_{t=0}^{\ell} \|R_t^{(\ell)}\|.$$

9.2.2. Norm bounds. It then remains to bound the different quantities in the lemma above; this is done in another section, using a trace bound method. The results are as follows:

Proposition 11. *On an event with probability $1 - c_0/\log(n)$, for any $k \leq c_1 \log(n)$, the following bounds hold with probability at least $1 - 1/\ln(n)^2$:*

$$\|\Delta^{(k-1)}\| \leq c d^3 \log(n)^{17} \left(\sqrt{\rho} \vee L\right)^k, \tag{9.6}$$

$$\|MB^{k-1}\| \leq \frac{c d^{7/2} L \log(n)^7 d^k L^k}{\sqrt{n}}, \tag{9.7}$$

$$\|\Delta^{(t-1)} \widetilde{M} B^{k-t-1}\| \leq \frac{c d^{13/2} L \log(n)^{24} d^k \left(\sqrt{\rho} \vee L\right)^k}{\sqrt{n}}, \tag{9.8}$$

$$\|R_t^{(k)}\| \leq \frac{c d^2 \log(n)^{22} d^k L^k}{n}. \tag{9.9}$$

Using these bounds, we are now finally able to prove (5.8):

Proof. By definition of ℓ, $d^\ell \leq n^{1/4}$ so most of the summands in Lemma 10 are negligible with respect to the others. More precisely, we have

$$\|B^\ell x\| \leq c_1 \left(\|\Delta^\ell\| + d \log(n) \sum_{t=1}^{\ell-1} \|\Delta^{(t-1)}\| \sum_{k=1}^{r} \left|\langle D_W \check{\chi}_k, B^{\ell-t-1} x\rangle\right|\right). \tag{9.10}$$

When $k \in [r_0]$, Lemma 9 implies that

$$\left|\langle D_W \check{\chi}_k, B^{\ell-t-1} x\rangle\right| \leq \sqrt{r} d^{3/2} L^2 \rho^{t/2} + \frac{c_4 b d^{3/2} \log(n)^{9/2} d^{2\ell} L^\ell}{n^{1/4}},$$

and by definition of ℓ, $d^{2\ell} L^\ell \leq (1 \wedge \sqrt{\rho})^\ell$ so the second term is bounded above by the first. On the other hand, for $k \in [r] \setminus [r_0]$, we can use equation (8.10) as follows:

$$\|(B^*)^t D_W \check{\chi}_i\|^2 \leq \mu_i^{2t+2} \Gamma_{V,ii}^{(t+1)} + \frac{c r b^2 d^4 L^2 \log(n)^6 d^{3t} L^{2t}}{\sqrt{n}}.$$

We now apply Lemma 3:

$$\Gamma_{V,ii}^{(t+1)} \leq \sum_{s=0}^{t+1} \frac{r d^2 L^2 \rho^s}{\mu_i^{2s}} \leq c r d^2 \log(n) L^2 \rho^{t+1} \mu_i^{-2t-2},$$

since $\mu_i^2 < \rho$; the second term being negligible before the first,

$$\left|\langle D_W \check{\chi}_k, B^{\ell-t-1} x\rangle\right| \leq \|(B^*)^{\ell-t-1} D_W \check{\chi}_i\| \leq c r d \log(n) L \rho^{(\ell-t)/2}.$$

We can now apply the above bounds on the scalar product as well as those of Proposition 11 to equation (9.10), and we get

$$\|B^\ell x\| \le c_2 d^{5/2} L \log(n)^{17} (\sqrt{\rho} \vee L)^\ell$$
$$+ c_3 r^2 d^6 L^2 \log(n)^{20} (\sqrt{\rho} \vee L)^\ell + c_4 d^4 \log(n)^{17} (\sqrt{\rho} \vee L)^\ell$$
$$\le c r^2 d^6 L^2 \log(n)^{20} (\sqrt{\rho} \vee L)^\ell,$$

which ends the proof of (5.8). ∎

10. Trace method: proof of Proposition 11

The aim of this section is to prove the bounds in Proposition 11; we leverage here the powerful trace method introduced by Füredi and Komlòs [20], and already used with success in [14] and [13]. We only prove (9.6) in this section, all other bounds being proven in the appendix.

Let m be a parameter to be fixed later. We start with the classical bound

$$\|\Delta^{(k-1)}\|^{2m} = \|\Delta^{(k-1)} \Delta^{(k-1)*}\|^m$$
$$= \|(\Delta^{(k-1)} \Delta^{(k-1)*})^m\| \le \mathrm{tr}\big((\Delta^{(k-1)} \Delta^{(k-1)*})^m\big).$$

Expanding the trace above gives

$$\|\Delta^{(k-1)}\|^{2m} \le \sum_{(e_1,\dots,e_{2m})} \prod_{i=1}^m (\Delta^{(k-1)})_{e_{2i-1},e_{2i}} (\Delta^{(k-1)})_{e_{2i+1},e_{2i}}$$
$$= \sum_{\gamma \in W_{k,m}} \prod_{i=1}^{2m} \underline{X}_{\gamma_{i,0} \gamma_{i,1}} \prod_{s=2}^k \underline{A}_{\gamma_{i,s-1} \gamma_{i,s}}, \tag{10.1}$$

where $W_{k,m}$ is the set of sequences of paths $(\gamma_1,\dots,\gamma_{2m})$ such that $\gamma_i = (\gamma_{i,0},\dots,\gamma_{i,k})$ is non-backtracking tangle-free of length k, and with boundary conditions that for all $i \in [m]$,

$$(\gamma_{2i,k-1}, \gamma_{2i,k}) = (\gamma_{2i-1,k-1}, \gamma_{2i-1,k}), \quad (\gamma_{2i+1,0}, \gamma_{2i+1,1}) = (\gamma_{2i,0}, \gamma_{2i,1}), \tag{10.2}$$

with the convention $\gamma_{2m+1} = \gamma_1$. All the random variables in the expression above are centered and independent as soon as they are supported by distinct edges, so the expectation of each term in the sum is zero except when each (unoriented) edge is visited at least twice. We let $W'_{k,m}$ be the set of all such sequences of paths. To $\gamma \in W'_{k,m}$, we associate the graph $G_\gamma = (V_\gamma, E_\gamma)$ of visited vertices and edges, and let

$$v_\gamma = |V_\gamma|, \quad e_\gamma = |E_\gamma|.$$

For an unoriented edge $e \in E_\gamma$, we define its multiplicity m_e as the number of times e is visited in γ; we also let S_γ be the set of starting edges in γ, that is

$$S_\gamma = \{(\gamma_{i,0}, \gamma_{i,1}) \mid i \in [2m]\}.$$

Using these definitions, we can bound the expectation as follows:

$$\mathbb{E}\big[\|\Delta^{(k-1)}\|^{2m}\big] \leq \sum_{\gamma \in W'_{k,m}} \prod_{e \in S_\gamma} \mathbb{E}\big[|\underline{X}_e| \cdot |\underline{A}_e|^{m_e - 1}\big] \prod_{e \notin S_\gamma} \mathbb{E}\big[|\underline{A}_e|^{m_e}\big].$$

We now bound the two terms in the products above: let e be an edge, and $p \geq 2$ be any multiplicity. Then conditioning on X_e,

$$
\begin{aligned}
\mathbb{E}\big[|\underline{A}_e|^{m_e}\big] &= P_e \mathbb{E}\big[|W_e - P_e\mathbb{E}[W_e]|^p\big] + (1 - P_e)P_e^p \mathbb{E}[W_e]^p \\
&\leq P_e L^{p-2}\Big(1 + \frac{d}{n}\Big)^{p-2} \mathbb{E}\big[(W_e - P_e\mathbb{E}[W_e])^2\big] + \Big(\frac{dL}{n}\Big)^{p-2} \frac{dP_e}{n} \mathbb{E}[W_e]^2 \\
&\leq P_e L^{p-2}\Big(1 + \frac{d}{n}\Big)^{p-2} \mathbb{E}[W_e^2] + P_e L^{p-2}\mathbb{E}[W_e^2]\Big(\frac{d}{n}\Big)^{p-2} \\
&\leq K_e L^{p-2}\Big(1 + \frac{d}{n}\Big)^p.
\end{aligned}
$$

The other product is trickier; whenever $p \geq 3$, a similar computation yields

$$\mathbb{E}\big[|\underline{X}_e| \cdot |\underline{A}_e|^{p-1}\big] \leq K_e L^{p-3}\Big(1 + \frac{d}{n}\Big)^p.$$

On the other hand if $p = 2$,

$$\mathbb{E}\big[|\underline{X}_e| \cdot |\underline{A}_e|\big] \leq \frac{d}{n} L \Big(1 + \frac{d}{n}\Big)^2.$$

As a consequence, for $\gamma \in W'_{k,m}$, we define $S'_\gamma \subseteq S_\gamma$ the set of starting edges with multiplicity 2. Then

$$\mathbb{E}\big[\|\Delta^{(k-1)}\|^{2m}\big] \leq \sum_{\gamma \in W'_{k,m}} \Big(1 + \frac{d}{n}\Big)^{2km} \Big(\frac{d}{n}\Big)^{|S'_\gamma|} d^{2m} L^{2km - 2e_\gamma} \prod_{e \notin S'_\gamma} K_e,$$

where we used $L^{-1} \leq d$ and $S_\gamma = 2m$.

We now partition the paths in $W'_{k,m}$ as follows: we say that $\gamma \sim \gamma'$ if there exists a permutation $\sigma \in \mathfrak{S}_n$ such that $\gamma_{i,t} = \sigma(\gamma'_{i,t})$ for all $i, t \in [2m] \times [k]$. Clearly, all parameters such as v_γ, e_γ and $|S'_\gamma|$ are constant on any equivalence class; therefore it makes sense to define $\mathcal{W}_{k,m}(v, e)$ the set of equivalence classes of $W'(k, m)$ such that $v_\gamma = v$ and $e_\gamma = e$. Then, a path counting argument performed in [14] yields the following estimation:

Lemma 11. *Let v, e be integers such that $e - v + 1 \geq 0$. Then*

$$\mathcal{W}_{k,m}(v, e) \leq k^{2m}(2km)^{6m(e-v+1)}. \tag{10.3}$$

All that remains to bound the sum above is to control the contribution of a single equivalence class; this is done through this lemma:

Lemma 12. *Let $\gamma \in W'_{k,m}$ such that $v_\gamma = v$, $e_\gamma = e$ and $|S'_\gamma| = s$. We have*

$$\sum_{\gamma' \sim \gamma} \prod_{f \notin S'_{\gamma'}} K_f \leq d^{2m} n^{v-e+s} \rho^e (\Psi^2)^{3(e-v)+8m}. \tag{10.4}$$

Proof. For a sequence of paths $\gamma \in W'_{k,m}$, denote by E'_γ the set $E_\gamma \setminus S'_\gamma$. Then, due to the boundary conditions in (10.2), the graph G'_γ induced by E'_γ is connected. We let v_j (resp. $v_{\geq j}$) be the number of vertices with degree j (resp. at least j) in G'_γ. Again, by (10.2), removing an edge in S'_γ does not create a vertex of degree 1; therefore we have

$$v_1 \leq 4m,$$

since a vertex of G_γ can only be of degree 1 if it is an endpoint of γ_i for some $i \in [2m]$. Additionally, edge and vertex counting yields

$$v_1 + v_2 + v_{\geq 3} \geq v - s, \quad v_1 + 2v_2 + 3v_{\geq 3} \leq 2(e - s),$$

since removing an edge in S'_γ removes at most one vertex from G_γ. Combining those inequalities gives

$$v_{\geq 3} + v_1 \leq 2(e - s) - 2(v - s) + 2v_1 \leq 2(e - v) + 8m; \tag{10.5}$$

this inequality encodes the fact that in a union of paths most vertices are of degree 2. We now reduce G'_γ into a multigraph $\widehat{G}_\gamma = (\widehat{V}_\gamma, \widehat{E}_\gamma)$ as follows: $\widehat{V}_\gamma$ is the set of vertices in G'_γ with degree different from 2, and we add an edge between two vertices x_1 and x_2 of $\widehat{V}_\gamma$ for each path between x_1 and x_2 in G'_γ. For $\widehat{f} \in \widehat{E}_\gamma$, we annotate $\widehat{f}$ with the length $q_{\widehat{f}}$ of its corresponding path in G'_γ.

We let $\widehat{v}$ and $\widehat{e}$ be the number of vertices and edges of $\widehat{G}_\gamma$; a sequence $\gamma' \sim \gamma$ is uniquely determined by an embedding of $\widehat{V}_\gamma$ in $[n]$ and for each edge $\widehat{f} \in \widehat{E}_\gamma$, an embedding of $\widehat{f}$ as a path of length $q_{\widehat{f}}$. As a result, we have

$$\sum_{\gamma' \sim \gamma} \prod_{f \notin S'_{\gamma'}} \varrho_f \leq \sum_{y_1, \ldots, y_{\widehat{v}} \in [n]^{\widehat{v}}} \prod_{\widehat{f} = (y_i, y_j) \in \widehat{E}_\gamma} \sum_{x_1, \ldots, x_{q_{\widehat{f}}-1} \in [n]} \prod_{t=1}^{q_{\widehat{f}}} K_{x_{t-1}, x_t}$$

$$= \sum_{y_1, \ldots, y_{\widehat{v}} \in [n]^{\widehat{v}}} \prod_{\widehat{f} = (y_i, y_j) \in \widehat{E}_\gamma} (K^{q_{\widehat{f}}})_{y_i, y_j}$$

$$\leq \sum_{y_1,\dots,y_{\hat{v}} \in [n]^{\hat{v}}} \prod_{\hat{f} \in \hat{E}_\gamma} \left(\frac{\Psi^2}{n} \rho^{q_{\hat{f}}}\right),$$

using (6.3) and recalling that $\Psi = L^2/\rho$. Now, notice that

$$\sum_{\hat{f} \in \hat{E}_\gamma} q_{\hat{f}} = |E'_\gamma| = e - s, \quad \hat{e} - \hat{v} = |E'_\gamma| - |V'_\gamma| \geq e - v - s;$$

further $\hat{e} \leq \hat{v} + e - v - s \leq 3(e - v) + 8m - s$ using (10.5) and the inequality above. We finally find

$$\sum_{\gamma' \sim \gamma} \prod_{f \notin S'_{\gamma'}} Q_f \leq n^{\hat{v} - \hat{e}} (\Psi^2)^{\hat{e}} \rho^{e-s}$$

$$\leq n^{v-e+s} \rho^{e-s} (\Psi^2)^{3(e-v)+8m-s},$$

which ends the proof of Lemma 12, since $\Psi^2 \geq 1$ and $\rho^{-1} \leq a$. ∎

We now are able to conclude this part; the contribution of one equivalence class in $\mathcal{W}_{k,m}(v, e)$ is less than

$$C_\gamma = \left(1 + \frac{d}{n}\right)^{2km} \left(\frac{d}{n}\right)^{|S'_\gamma|} d^{2m} L^{2km-2e} \sum_{\gamma' \sim \gamma} \prod_{e \notin S'_{\gamma'}} K_e$$

$$\leq c_1^{2m} d^{6m} n^{-|S'_\gamma|} L^{2km-2e} n^{v-e+|S'_\gamma|} \rho^e (\Psi^2)^{3(e-v)+8m}$$

$$\leq c_1^{2m} d^{6m} n^{v-e} (\rho\Psi)^{km-e} \rho^e (\Psi^2)^{3(e-v)+8m}$$

$$\leq c_1^{2m} d^{6m} \rho^{km} \left(\frac{\Psi}{d}\right)^{km-e} n^{1-g} (\Psi^2)^{3g+8m},$$

with $g = e - v + 1$ and we used that $L = \sqrt{\rho\Psi/d}$ and the bound

$$\left(1 + \frac{d}{n}\right)^k \leq \exp\left(\frac{dk}{n}\right) \leq c_1.$$

Summing over all equivalence classes now gives

$$\mathbb{E}\big[\|\Delta^{(k-1)}\|^{2m}\big] \leq \sum_{e=1}^{km} \sum_{v=1}^{e+1} |\mathcal{W}_{k,m}(v, e)| \max_{[\gamma] \in \mathcal{W}_{k,m}(v,e)} C_\gamma$$

$$\leq \sum_{e=1}^{km} \sum_{v=1}^{e+1} k^{2m} (2km)^{6m(e-v+1)} c_1^{2m} d^{6m} \rho^{km} \left(\frac{\Psi}{d}\right)^{km-e} n^{1-g} (\Psi^2)^{3g+8m}$$

$$\leq n(c_1 d^3 k)^{2m} \rho^{km} \sum_{e=1}^{km} \left(\frac{\Psi}{d}\right)^{km-e} \sum_{g=0}^{\infty} \left(\frac{\Psi^6 (2km)^{6m}}{n}\right)^g. \tag{10.6}$$

We set the parameter m to

$$m = \left\lceil \frac{\log(n/\Psi^6)}{12\log(\log(n))} \right\rceil;$$

when $n \geq c_2 \Psi^6$ for some absolute constant c_2, we have

$$\frac{\Psi^6(2km)^{6m}}{n} < \frac{1}{2} \quad n^{1/2m} \leq \log(n)^{12}.$$

The infinite sum inside (10.6) thus converges, and

$$\mathbb{E}\big[\|\Delta^{(k-1)}\|^{2m}\big]^{1/2m} \leq c_3 d^3 \log(n)^{14} \sqrt{\rho}^k \left(1 \vee \sqrt{\frac{\Psi}{d}}\right)^k.$$

Finally, from the definition of Ψ, we have $\sqrt{\rho}(1 \vee \sqrt{\Psi/d}) = \sqrt{\rho} \vee L$, hence (9.6) by a Markov bound.

A. Applications of Theorem 2

A.1. Proof of Proposition 1

Let x be an eigenvector of B associated with the eigenvalue λ; the eigenvalue equation for x reads

$$\lambda x_e = \sum_{e \to f} W_f x_f. \tag{A.1}$$

On the other hand, the definition $y = S^* D_W x$ expands to

$$y_i = \sum_{e:e_1=i} W_e x_e.$$

Applying equation (A.1) to e and e^{-1} yields

$$\lambda x_e = y_{e_2} - W_e x_{e^{-1}}, \quad \lambda x_{e^{-1}} = y_{e_1} - W_e x_e,$$

and as a result

$$\lambda^2 x_e = \lambda y_{e_2} - \lambda W_e x_{e^{-1}} = \lambda y_{e_2} - W_e(y_{e_1} - W_e x_e).$$

Rearranging the terms, we find an expression for x_e:

$$x_e = \frac{\lambda y_{e_2} - W_e y_{e_1}}{\lambda^2 - W_e^2}; \tag{A.2}$$

in particular $y \neq 0$ if $x \neq 0$. Plugging (A.2) into the eigenvalue equation (A.1), for $i, j \in [n]$, we get

$$\frac{\lambda^2 y_i - \lambda W_{ij} y_j}{\lambda^2 - W_{ij}^2} = \sum_{\substack{k \sim i \\ k \neq j}} W_{ik} \frac{\lambda y_k - W_{ik} y_i}{\lambda^2 - W_{ik}^2},$$

and we rearrange to find

$$\frac{\lambda^2 y_i}{\lambda^2 - W_{ij}^2} - \frac{W_{ij}^2 y_i}{\lambda^2 - W_{ij}2} = \sum_{k \sim i} \frac{\lambda W_{ik}}{\lambda^2 - W_{ik}^2} y_k - \sum_{k \sim i} \frac{W_{ik}^2}{\lambda^2 - W_{ik}^2} y_i.$$

The fraction on the left-hand side cancels out, and writing the right-hand side as a matrix product

$$y = \widetilde{A}(\lambda) y - \widetilde{D}(\lambda) y,$$

the desired result.

A.2. Proof of Theorem 3

Our first step is to show that the matrices involved in Proposition 1 approximate the matrices A and ρI. If $\lambda^2 \geq 2L^2$, we have

$$|\lambda \widetilde{A}_{ij}(\lambda) - A_{ij}| = \mathbf{1}\{i \sim j\} \left| \frac{W_{ij}}{1 - (W_{ij}^2/\lambda^2)} - W_{ij} \right| \leq \mathbf{1}(i \sim j) \frac{2L W_{ij}^2}{\lambda^2},$$

which implies using the Gershgorin circle theorem

$$\left\| \lambda \widetilde{A}_{ij}(\lambda) - A_{ij} \right\| \leq \frac{2L}{\lambda^2} \max_i \sum_{j \sim i} W_{ij}^2 \leq \frac{4L\rho}{\lambda^2}. \tag{A.3}$$

Similarly,

$$\left| \lambda^2 \widetilde{D}_{ii}(\lambda) - \rho \right| \leq \frac{2L^2}{\lambda^2} \sum_{j \sim i} W_{ij}^2 + \left| \sum_{j \sim i} W_{ij}^2 - \rho \right|$$

$$\leq \left(\frac{4L^2}{\lambda^2} + \varepsilon \right) \cdot \rho. \tag{A.4}$$

We now take $\lambda = \lambda_i$ with $i \in [r_0]$; then there is a vector y that is a singular value of

$$-\lambda_i \Delta(\lambda_i) = A - \left(\lambda_i + \frac{\rho}{\lambda_i} \right) I + \left(\lambda \widetilde{A}(\lambda_i) - A \right) - \lambda_i^{-1} \left(\lambda^2 \widetilde{D}(\lambda) - \rho I \right).$$

We can thus apply Weyl's inequality [34] to find that there exists an eigenvalue ν_i of A such that

$$\left| \nu_i - \left(\lambda_i + \frac{\rho}{\lambda_i} \right) \right| \leq \frac{4L\rho}{\lambda_i^2} + \left(\frac{4L^2}{\lambda_i^2} + \varepsilon \right) \cdot \frac{\rho}{\lambda_i}.$$

Now, we use Theorem 2 to find that $|\lambda_i - \mu_i| \leq \sigma$, and we have $\sigma = o(\rho)$ whenever n is large enough by virtue of (6.1). Since

$$|\lambda_i - \mu_i| \leq \sigma, \quad \left| \frac{\rho}{\lambda_i} - \frac{\rho}{\mu_i} \right| \leq \frac{\rho}{\lambda_i \mu_i} \sigma \leq c_0 \sigma,$$

equation (3.2) ensues by noticing that $\lambda_i > c_1 \mu_i$ for some constant c_1 and σ is negligible before the other error terms.

Assume now that $\delta_i \geq 2\sigma$; examining the proof of Theorem 2, we have the existence of an eigenvector ξ of B associated with λ_i such that

$$\|\xi - u_i\| \leq \frac{3\sigma \|u_i\|}{\delta_i - \sigma}.$$

Proposition 1 implies that the vector $y = S^* D_W \xi$ is a null vector of the deformed laplacian $\Delta(\lambda) = I - \widetilde{A}(\lambda) + \widetilde{D}(\lambda)$. Notice that the matrix $S^* D_W^2 S$ is a diagonal matrix such that

$$[S^* D_W^2 S]_{ii} = \sum_{j \sim i} W_{ij}^2 \leq 2\rho,$$

from which we have

$$\|y - S^* D_W u_i\| \leq \frac{6\sigma \sqrt{\rho} \|u_i\|}{\delta_i - \sigma}.$$

We now follow the line of proof of Theorem 2; we first find

$$\langle S^* D_W u_i, \varphi_i \rangle = \mu_i^{-\ell} \langle B^\ell \chi_i, D_W \check{\chi}_i \rangle,$$

and combine it with (8.8) to obtain

$$\left| \langle S^* D_W u_i, \varphi_i \rangle - \mu_i \right| \leq \sigma. \tag{A.5}$$

Computing $\|S^* D_W u_i\|$ is trickier; we find

$$\begin{aligned}
\langle S^* D_W u_i, S^* D_W u_i \rangle &= \mu_i^{-2\ell} \langle S^* D_W B^\ell \chi_i, S^* D_W B^\ell \chi_i \rangle \\
&= \mu_i^{-2\ell} \langle S^* D_W B^\ell \chi_i, T^* J D_W B^\ell \chi_i \rangle \\
&= \mu_i^{-2\ell} \langle T S^* D_W B^\ell \chi_i, (B^*)^\ell D_W \check{\chi}_i \rangle.
\end{aligned}$$

Writing the coefficients of $T S^* D_W$ explicitly, we have

$$[T S^* D_W]_{ef} = W_f \sum_{i \in [n]} \mathbf{1}\{e_2 = i\} \mathbf{1}\{f_1 = i\} = B_{ef} + [J D_W]_{ef},$$

which yields

$$\langle S^* D_W u_i, S^* D_W u_i \rangle = \mu_i^{-2\ell} \left(\langle B^{2\ell+1} \chi_i, D_W \check{\chi}_i \rangle + \langle B^\ell D_W \check{\chi}_i, B^\ell D_W \check{\chi}_i \rangle \right).$$

Those scalar products correspond to equations (8.8) and (8.10), respectively, and we thus get

$$\left|\|S^*D_W u_i\|^2 - \mu_i^2\big(1 + \Gamma_{V,ii}^{(\ell)}\big)\right| \le 2\sigma.$$

The hypothesis $K\mathbf{1} = \rho\mathbf{1}$ allows us to approximate $\Gamma_{V,ii}^{(\ell)}$ efficiently:

$$\Gamma_{V,ii}^{(\ell)} = \sum_{t=0}^{\ell} \frac{\langle \mathbf{1}, K^{t+1}\varphi^{i,i}\rangle}{\mu_i^{2t+2}} = \sum_{t=0}^{\ell} \left(\frac{\rho}{\mu_i^2}\right)^{t+1}$$

since $\|\varphi_i\| = 1$, and we have as in the proof of Theorem 2

$$\left|\Gamma_{V,ii}^{(\ell)} - \frac{\rho/\mu_i^2}{1 - \rho/\mu_i^2}\right| \le \sigma.$$

Gathering the previous bounds, we eventually arrive at

$$\left|\|S^*D_W u_i\|^2 - \frac{\mu_i^2}{1 - \rho/\mu_i^2}\right| \le 3\sigma. \tag{A.6}$$

The exact same computations imply that

$$\|u_i\|^2 \le \frac{d}{1 - \rho/\mu_i^2} + c_5\sigma,$$

and thus noticing that $\mu_i \ge \sqrt{\rho}$,

$$\left\|\frac{y}{\|y\|} - \frac{S^*D_W u_i}{\|S^*D_W u_i\|}\right\| \le \frac{c_6\,\sigma\sqrt{d}}{\delta_i - \sigma}.$$

Combining this error bound with (A.5) and (A.6), we find the following result:

$$\left|\left\langle \frac{y}{\|y\|}, \varphi_i\right\rangle - \sqrt{1 - \frac{\rho}{\mu_i^2}}\right| \le \frac{c_7\,\sigma\sqrt{d}}{\delta_i - \sigma}.$$

The final step is to use the Davis–Kahan theorem [35] as follows: there exists an eigenvector ζ of A with associated eigenvalue ν_i, and such that

$$\left\|\zeta - \frac{y}{\|y\|}\right\| \le \frac{c_8\big(\frac{4L\rho}{\lambda_i^2} + \big(\frac{4L^2}{\lambda_i^2} + \varepsilon\big)\cdot\frac{\rho}{\lambda_i}\big)}{\delta_i}.$$

This error term dominates all the other ones found above, hence the bound in Theorem 3.

The proof of Corollary 1 follows along the same lines; however, we have directly

$$\widetilde{A}(\lambda) = \frac{\lambda A}{\lambda^2 - 1}, \quad \widetilde{D}(\lambda) = \frac{d_0}{\lambda^2 - 1}I,$$

and thus the approximation bounds (A.3) and (A.4) become superfluous.

A.3. Proof of Theorem 4

We first link the SBM setting to the one of Theorem 2. In the unweighted case, we have $Q = K = P$, and the eigenvector equation $P\mathbf{1} = \alpha\mathbf{1}$ yields $\rho = \alpha$. It is easy to check that whenever n is large enough, the r_0 defined in Theorem 4 satisfies the assumptions of Theorem 2, with $\tau = 1/(\alpha\mu_{r_0}^2) < 1$. Equation (3.4) ensures that $\|\varphi_i\|_\infty \leq c/\sqrt{n}$ for some absolute constant $c > 0$, therefore $b = O(1)$. Finally, since $\tau^{-1} = \alpha\mu_{r_0}$, we have

$$C_0 \leq c\alpha \log{(n)}^{25}, \quad n_0 \leq \exp\big(c\log(d)\log(\log(n))\big).$$

An application of Theorem 2 thus directly yields the bound on the eigenvalues of B; regarding the eigenvectors, notice that as in the proof of Theorem 3

$$\|u_i\|^2 = \frac{\alpha}{1 - 1/(\alpha\mu_i^2)} + O(\sigma), \quad \|T\varphi_i\| = \alpha + O(\sigma),$$

which gives

$$\langle \xi, \xi_i \rangle = \sqrt{1 - \frac{1}{\alpha\mu_i^2}} + O(\sigma).$$

A.4. Proof of Theorem 5 and Proposition 2

Letting again Θ be the $n \times 2$ group membership matrix, we find as in the proof of Theorem 4 that we have $Q = \Theta\tilde{Q}\Theta^*$ and $K = \Theta\tilde{K}\Theta^*$, with

$$\tilde{Q} = \frac{1}{2}\begin{pmatrix} a\mathbb{E}_\mathbb{P}[w] & b\mathbb{E}_\mathbb{Q}[w] \\ b\mathbb{E}_\mathbb{Q}[w] & a\mathbb{E}_\mathbb{P}[w] \end{pmatrix}, \quad \tilde{K} = \frac{1}{2}\begin{pmatrix} a\mathbb{E}_\mathbb{P}[w^2] & b\mathbb{E}_\mathbb{Q}[w^2] \\ b\mathbb{E}_\mathbb{Q}[w^2] & a\mathbb{E}_\mathbb{P}[w^2] \end{pmatrix}.$$

This implies first that

$$\rho = \frac{a\mathbb{E}_\mathbb{P}[w^2] + b\mathbb{E}_\mathbb{Q}[w^2]}{2},$$

and that the vector $\Theta\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is an eigenvector of Q associated with the eigenvalue

$$\mu_2 = \frac{a\mathbb{E}_\mathbb{P}[w] - b\mathbb{E}_\mathbb{Q}[w]}{2}.$$

All other hypotheses of Theorem 2 are easy to check, and we find that the announced results hold as soon as $\mu_2^2 > \rho \vee L$, or

$$\frac{(a\mathbb{E}_\mathbb{P}[w^2] + b\mathbb{E}_\mathbb{Q}[w^2]) \vee L}{(a\mathbb{E}_\mathbb{P}[w] - b\mathbb{E}_\mathbb{Q}[w])^2} < 1.$$

Now, let us disregard for a moment the condition on L, and compute ρ:

$$\rho = \frac{1}{2}\int_\mathscr{L}\big(af(\ell) + bg(\ell)\big)w(\ell)^2\,\mathrm{d}m(\ell).$$

Define a scalar product on $\ell^\infty(\mathscr{L})$, the set of all bounded functions from $\mathscr{L}$ to $\mathbb{R}$, as

$$\langle h_1, h_2 \rangle_{\mathscr{L}} = \int_{\mathscr{L}} \big(af(\ell) + bg(\ell)\big) h_1(\ell) h_2(\ell) \, dm(\ell);$$

then $\rho = \|w\|_{\mathscr{L}}^2$, and applying the Cauchy–Schwarz theorem

$$\rho \cdot \left\| \frac{af - bg}{af + bg} \right\|_{\mathscr{L}}^2 \geq \left\langle w, \frac{af - bg}{af + bg} \right\rangle_{\mathscr{L}}^2 = \mu_2^2.$$

This implies that the signal-to-noise ratio μ_2^2/ρ is maximized whenever

$$w(\ell) = \frac{af(\ell) - bg(\ell)}{af(\ell) + bg(\ell)},$$

and in this case

$$\beta = \frac{\mu_2^2}{\rho} = \frac{1}{2} \int_{\mathscr{L}} \frac{(af(\ell) - bg(\ell))^2}{af(\ell) + bg(\ell)} \, dm(\ell).$$

In particular, we have $\mu_2 = \rho = \beta$, so $\beta > 1$ implies $\mu_2 > 1$. It remains to notice that $w(\ell) \leq 1$ for any ℓ, so the condition $\mu_2 \geq L$ is redundant as assumed.

A.5. Proof of Theorem 6

For $i, j \in [n]$, we note $W_{ij} = m_{ij} + s_{ij} Z_{ij}$ with $Z \sim \mathcal{N}(0, 1)$ a standard gaussian random variable. Let $\tilde{L} = 2\sqrt{\log(n)}$; a well-known tail bound for gaussians reads

$$\mathbb{P}\big(|Z_{ij}| \geq \tilde{L}\big) \leq \frac{2}{\tilde{L}} e^{-\tilde{L}/2} \leq \frac{1}{n^2 \sqrt{\log(n)}}. \tag{A.7}$$

We now define the modified matrix $\tilde{W}$ with

$$\tilde{W}_{ij} = m_{ij} + s_{ij} Z_{ij} \mathbf{1}\{|Z_{ij}| \leq \tilde{L}\},$$

with $\tilde{Q}$ and $\tilde{K}$ the associated expected and variance matrices. It is readily seen that $\tilde{Q} = Q$, and that the variables $\tilde{W}_{ij}$ are bounded by

$$L = \sup_{i,j} |m_{ij}| + \tilde{L} \sup_{i,j} s_{ij}.$$

By a union bound, we have

$$\mathbb{P}(\tilde{W} \neq W) = \mathbb{P}\big(Z_{ij} > \tilde{L} \text{ for some } i \in [n]\big) \leq \binom{n}{2} \frac{1}{n^2 \sqrt{\log(n)}} \leq \frac{1}{2\sqrt{\log(n)}},$$

and whenever $\widetilde{W} = W$, then the modified non-backtracking matrix coincides with the original one. Finally, notice that for $i, j \in [n]$,

$$\operatorname{Var}\big(Z_{ij}\mathbf{1}\{|Z_{ij}| \leq \widetilde{L}\}\big) \leq 1,$$

which implies using the Perron–Frobenius theorem that $\rho(\widetilde{K}) \leq \rho(K)$. Theorem 2 then applies to the modified couple $(P, \widetilde{W})$ and the announced result follows.

B. Computing functionals on trees

We prove in this section the martingale estimates of Proposition 6 and Proposition 7.

B.1. Study of compound Poisson processes

Many proofs in this section rely on computations of Poisson compound processes, i.e. Poisson sums of random variables. For convenience, we gather them all in the following lemma:

Lemma 13. *Let N be a* $\operatorname{Poi}(d)$ *random variable, and (X_i), (Y_i), (Z_i) three i.i.d. sequences of random variables, independent from N, such that X_i and Y_j (resp. Y_i and Z_j, or Z_i and X_j) are independent whenever $i \neq j$. Denote by A, B the random variables*

$$A = \sum_{i=1}^{N} X_i, \quad B = \sum_{i=1}^{N} Y_i.$$

Then the following identities hold:

$$\mathbb{E}[A] = d\,\mathbb{E}[X], \quad \mathbb{E}[B] = d\,\mathbb{E}[Y], \tag{B.1}$$

$$\mathbb{E}[AB] = d\,\mathbb{E}[XY] + d^2\mathbb{E}[X]\mathbb{E}[Y] = d\,\mathbb{E}[XY] + \mathbb{E}[A]\mathbb{E}[B], \tag{B.2}$$

$$\mathbb{E}\left[\sum_{i=1}^{N} Z_i\left(\sum_{j\neq i} X_j\right)\right] = d\,\mathbb{E}[A]\mathbb{E}[Z], \tag{B.3}$$

$$\mathbb{E}\left[\sum_{i=1}^{N} Z_i\left(\sum_{j\neq i} X_j\right)\left(\sum_{k\neq i} Y_k\right)\right] = d\,\mathbb{E}[AB]\mathbb{E}[Z]. \tag{B.4}$$

Although the first two identities are well known, we provide a full proof of this lemma:

Proof. The sequence (X_i) being independent from N, we immediately find that

$$\mathbb{E}[A \mid N] = N\mathbb{E}[X],$$

from which equation (B.1) is derived. We then write

$$AB = \left(\sum_{i=1}^{N} X_i\right)\left(\sum_{i=1}^{N} Y_i\right) = \sum_{i=1}^{N} X_i Y_i + \sum_{i \neq j} X_i Y_j,$$

and using the independence property of $(X_i)_i$ and $(Y_i)_i$ yields

$$\mathbb{E}[AB \mid N] = N\mathbb{E}[XY] + N(N-1)\mathbb{E}[X]\mathbb{E}[Y].$$

Since N is a Poisson random variable, $\mathbb{E}[N(N-1)] = d^2$, hence (B.2).

We now move onto the third equation; rearranging terms gives

$$\sum_{i=1}^{N} Z_i \left(\sum_{j \neq i} X_j\right) = \sum_{i \neq j} Z_i X_j,$$

and therefore the conditional expectation given N is $N(N-1)\mathbb{E}[X]\mathbb{E}[Z]$. Using again that $\mathbb{E}[N(N-1)] = d^2$ brings (B.3).

Similarly, we can rearrange

$$\sum_{i=1}^{N} Z_i \left(\sum_{j \neq i} X_j\right)\left(\sum_{k \neq i} Y_k\right) = \sum_{j \neq i} X_j Y_j Z_i + \sum_{i \neq j \neq k} X_i Y_j Z_j,$$

and take conditional expectations on both sides to arrive at

$$\mathbb{E}\left[\sum_{i=1}^{N} Z_i \left(\sum_{j \neq i} X_j\right)\left(\sum_{k \neq i} Y_k\right) \mid N\right]$$
$$= N(N-1)\mathbb{E}[XY]\mathbb{E}[Z] + N(N-1)(N-2)\mathbb{E}[X]\mathbb{E}[Y]\mathbb{E}[Z].$$

Again, the expected value of $N(N-1)(N-2)$ is d^3, and we finally find

$$\mathbb{E}\left[\sum_{i=1}^{N} Z_i \left(\sum_{j \neq i} X_j\right)\left(\sum_{k \neq i} Y_k\right)\right] = d^2\mathbb{E}[XY]\mathbb{E}[Z] + d^3\mathbb{E}[X]\mathbb{E}[Y]\mathbb{E}[Z]$$
$$= d\mathbb{E}[AB]\mathbb{E}[Z],$$

which ends the proof. ∎

B.2. Decomposing the tree functionals

We now fix $t \geq 1$, $x \in [n]$ and two vectors $\varphi, \varphi' \in \mathbb{R}^n$ for the rest of the section. Let N be the number of children of the root of T, and $(T_k, I_k)_{k \leq N}$ the subtrees at depth 1. We further introduce the following first moment notations:

$$g_\varphi(t, x) = \mathbb{E}\left[f_{\varphi,t}(T_x, x)\right], \quad h_{\varphi,\varphi'}(t, x) = \mathbb{E}\left[f_{\varphi,t}(T_x, x)f_{\varphi',t}(T_x, x)\right].$$

We begin by a small elementary computation: let $\phi \in \mathbb{R}^n$ be any vector. Then,

$$\mathbb{E}\big[W_{xI_k}\phi(I_k)\big] = \sum_{y\in[n]} \frac{P_{xy}}{d_x} \mathbb{E}[W_{xy}]\phi(y) = \frac{[Q\phi](x)}{d_x}. \tag{B.5}$$

Now, by linearity, we have

$$f_{\varphi,t}(T_x, x) = \sum_{k=1}^{N} W_{xI_k} f_{\varphi,t-1}(T_k, I_k). \tag{B.6}$$

By definition of the Galton–Watson tree, the random variables

$$X_k = W_{xI_k} f_{\varphi,t-1}(T_k, I_k), \quad Y_k = W_{xI_k} f_{\varphi',t-1}(T_k, I_k)$$

satisfy the assumptions of Lemma 13. Furthermore, conditioning on the value of I_k, we can compute $\mathbb{E}X_k$:

$$\mathbb{E}\big[W_{xI_k} f_{\varphi,t-1}(T_k, I_k)\big] = \mathbb{E}\big[W_{xI_k} g_\varphi(t-1, I_k)\big]$$
$$= \frac{[Qg_\varphi(t-1, \cdot)](x)}{d_x}.$$

Applying (B.1), and from the definition of g_φ, we come to the following recurrence relation:

$$g_\varphi(t, x) = \big[Qg(t-1, \cdot)\big](x).$$

Solving this recurrence is straightforward, and we find

$$g_\varphi(t, \cdot) = Q^t g_\varphi(0, \cdot) = Q^t\varphi,$$

which implies (8.1).

Using now equation (B.2) from Lemma 13, we derive

$$h_{\varphi,\varphi'}(t, x) = d_x \mathbb{E}\big[W_{xI_k}^2 h_{\varphi,\varphi'}(t-1, I_k)\big] + g_\varphi(t, x)g_{\varphi'}(t, x)$$
$$= \big[Kh_{\varphi,\varphi'}(t, \cdot)\big](x) + g_\varphi(t, x)g_{\varphi'}(t, x), \tag{B.7}$$

from which we can solve for $h_{\varphi,\varphi'}$:

$$h_{\varphi,\varphi'}(t, \cdot) = Kh_{\varphi,\varphi'}(t-1, \cdot) + (Q^t\varphi) \odot (Q^t\varphi')$$
$$= \sum_{s=0}^{t} K^s\big[(Q^{t-s}\varphi) \odot (Q^{t-s}\varphi')\big].$$

The eigenvector equations for φ_i and φ_j then imply (8.2).

Consider now the function $F_{i,t}(T_x, x) = f_{\varphi_i, t+1}(T_x, x) - \mu_i f_{\varphi_i, t}(T_x, x)$, and its associated first moment functions

$$G(t, x) = \mathbb{E}\big[F_{i,t}(T_x, x)\big], \quad H(t, x) = \mathbb{E}\big[F_{i,t}(T_x, x)^2\big].$$

The linearity of $f_{\varphi_i, t}$ implies that $F_{i,t}$ also verifies equation (B.6), and therefore

$$G(t, \cdot) = Q^t G(0, \cdot) = 0$$

for all $t \geq 0$. Equation (B.7) thus reduces to

$$\begin{aligned}
H(t, \cdot) &= K^t H(0, \cdot) \\
&= K^t \big(h_{\varphi_i, \varphi_i}(1, \cdot) - \mu_i^2 \varphi^{i,i}\big) \\
&= K^t \big(\mu_i^2 \varphi^{i,i} + K \varphi^{i,i} - \mu_i^2 \varphi^{i,i}\big) \\
&= K^{t+1} \varphi^{i,i},
\end{aligned}$$

which ends the proof.

B.3. Edge functionals

Most of the handiwork needed to prove Proposition 7 was done in Lemma 13; indeed, in the tree (T_x, x), the edge transformation on $f_{\varphi, t}$ can be written as

$$\vec{\partial}_w f_{\varphi, t}(T_x, x) = \sum_{j=1}^{N} w_{I_j} \sum_{k \neq j} W_{x I_k} f_{\varphi, t-1}\big(T(k), I_k\big).$$

We define accordingly the random variables

$$X_k = W_{x I_k} f_{\varphi, t-1}\big(T(k), I_k\big), \quad Y_k = W_{x I_k} f_{\varphi', t-1}\big(T(k), I_k\big), \quad Z_k = w_{I_k},$$

that verify the assumptions of Lemma 13. Computing $\mathbb{E}[Z]$ is straightforward:

$$\mathbb{E}[Z] = \sum_{y \in [n]} \frac{P_{xy}}{d_x} \mathbb{E}[w_y] = [P \overline{w}](x).$$

Hence, we can apply equation (B.3) to those variables, to deduce (8.4). Similarly, the product transformation has the form

$$\begin{aligned}
\vec{\partial}_w (f_{\varphi, t} \cdot f_{\varphi', t})(T_x, x) = \sum_{j=1}^{N} w_{I_j} \Bigg(&\sum_{k \neq j} W_{x I_k} f_{\varphi, t-1}\big(T(k), I_k\big) \Bigg) \\
&\times \Bigg(\sum_{k \neq j} W_{x I_k} f_{\varphi', t-1}\big(T(k), I_k\big) \Bigg),
\end{aligned}$$

which using (B.4) implies (8.5). Finally, equation (8.6) is proved with the exact same technique, considering $F_{i,t}(T_x, x)$ instead of $f_{\varphi, t}(T_x, x)$.

C. Near eigenvectors: computations

We finish here the proof of Proposition 10. First, let

$$f(g, o) = \mathbf{1}_{(g,o)_{t+1} \text{ has no cycles }} \varphi_j(o) f_{\varphi, t+1}(g, o).$$

Then f is $(t + 1)$-local, and we have

$$|f(g, o)| \leq \|\varphi_i\|_\infty \|\varphi_j\|_\infty |\partial(g, o)_{t+1}| L^{t+1}$$

$$\leq \frac{b^2}{n} |(g, o)|_{t+1} L^{t+1} := \psi(g, o).$$

On the other hand, the scalar product $\langle B^t \chi_i, D_W \check{\chi}_j \rangle$ can be written as

$$\langle B^t \chi_i, D_W \check{\chi}_j \rangle = \sum_{e \in \vec{E}} W_e \varphi_j(e_1) \sum_\gamma \prod_{s=1}^t W_{\gamma_s \gamma_{s+1}} \varphi_i(\gamma_{t+1})$$

$$= \sum_{e \in \vec{E}} \varphi_j(e_1) \sum_\gamma \prod_{s=0}^t W_{\gamma_s \gamma_{s+1}} \varphi_i(\gamma_{t+1}),$$

where the sum ranges over all non-backtracking paths $\gamma = (\gamma_0, \ldots, \gamma_{t+1})$ such that $(\gamma_0, \gamma_1) = e$. It follows that

$$\left| \langle B^t \chi_i, D_W \check{\chi}_j \rangle - \sum_{v \in V} f(G, v) \right| \leq \left| \sum_{e: e_1 \notin \mathcal{V}_{t+1}} [B^t \chi_i](e)[D_W \check{\chi}_j](e) \right|$$

$$\leq 2|\mathcal{V}_{t+1}| \max_v \psi(G, v),$$

using the tangle-free property as before. This time, the results from Section 7.2 yield

$$\max_v \psi(G, v) \leq \frac{c_1 b^2 \log(n)^2 d^{t+1} L^{t+1}}{n}, \quad \|\psi\|_\star \leq \frac{c_1 b^2 \log(n)^3 d^{t+1} L^{t+1}}{n},$$

and the expected value on the tree is

$$\sum_{x \in [n]} \mathbb{E}[f(T_x, x)] = \sum_x \varphi_j(x) \mu_i^{t+1} \varphi_i(x) = \mu_i^{t+1} \delta_{ij}.$$

Concluding,

$$\left| \langle B^t \chi_i, D_W \check{\chi}_j \rangle - \mu_i^{t+1} \delta_{ij} \right|$$

$$\leq \frac{c_2 b^2 \log(n)^4 d^{2t+3} L^{t+1}}{n} + \frac{c_3 r b^2 \log(n)^6 d^{2t+3} L^{t+1}}{\sqrt{n}}$$

$$\leq \frac{c_4 r b^2 d^3 L \log(n)^6 d^{2t} L^t}{\sqrt{n}},$$

which proves (8.8).

Now, let

$$f(g,o) = \mathbf{1}_{(g,o)_t \text{ has no cycles}} \, \vec{\partial}_{\mathbf{1}}\big[f_{\varphi_i,t} \cdot f_{\varphi_j,t} \big](g,o).$$

Again, f is t-local, and we have

$$|f(g,o)1| \leq \|\varphi_i\|_\infty \|\varphi_j\|_\infty \deg(o)|\partial(g,o)_t|^2 L^{2t}$$

$$\leq \frac{b^2}{n} \deg(o)|(g,o)_t|^2 L^{2t} := \psi(g,o).$$

By definition of the $\vec{\partial}$ operator, for $v \in V$, we have

$$f(g,v) = \sum_{e:e_2=v} [B^t \chi_i](e)[B^t \chi_j](e).$$

Hence,

$$\left| \langle B^t \chi_i, B^t \chi_j \rangle - \sum_{v \in V} f(G,v) \right| = \left| \sum_{e:e_2 \notin \mathcal{V}_t} [B^t \chi_i](e)[B^t \chi_j](e) \right|$$

$$\leq 2|\mathcal{V}_t| \max_v \psi(G,v),$$

using the tangle-free property as before. This time, the results from Section 7.2 yield

$$\max_v \psi(G,v) \leq \frac{c_5 \, b^2 \log(n)^3 d^{2t+1} L^{2t}}{n}, \quad \|\psi\|_\star \leq \frac{c_5 \, b^2 \log(n)^4 d^{2t+1} L^{2t}}{n},$$

and we can compute the expected value on the tree:

$$\sum_{x \in [n]} \mathbb{E}\big[f(T_x, x) \big] = \sum_x [P\mathbf{1}](x) \mu_i^t \mu_j^t \sum_{s=0}^t \frac{[K^s \varphi^{i,j}](x)}{(\mu_i \mu_j)^s}$$

$$= (\mu_i \mu_j)^t \sum_{s=0}^t \frac{\langle P\mathbf{1}, K^s \varphi^{i,j} \rangle}{(\mu_i \mu_j)^s} = (\mu_i \mu_j)^t \Gamma_{U,ij}^{(t)}.$$

Gathering those estimates, we find

$$\left| \langle B^t \chi_i, B^t \chi_j \rangle - (\mu_i \mu_j)^t \Gamma_{U,ij}^{(t)} \right|$$

$$\leq \frac{c_6 \, b^2 \log(n)^5 d^{3t+2} L^{2t}}{n} + \frac{c_7 \, rb^2 \log(n)^7 d^{3t+2} L^{2t}}{\sqrt{n}}$$

$$\leq \frac{c_8 \, rb^2 d^2 \log(n)^7 d^{3t} L^{2t}}{\sqrt{n}},$$

which proves (8.9).

Next is (8.10); we first notice that the parity-time equation (2.2) implies that

$$\langle (B^*)^t D_W \check{\chi}_i, (B^*)^t D_W \check{\chi}_j \rangle = \langle D_W B^t \chi_i, D_W B^t \chi_j \rangle.$$

Similarly to the previous computation, we therefore let $w_o = (W_{1o}^2, \ldots, W_{no}^2)$, and

$$f(g, o) = \mathbf{1}_{(g,o)_t \text{ has no cycles}} \vec{\partial}_{w_o} [f_{\varphi_i, t} f_{\varphi_j, t}](g, o).$$

We have similarly

$$|f(g, o)| \leq \frac{b^2}{n} |(g, o)_t|^2 L^{2t+2} := \psi(g, o),$$

Now,

$$\max_v \psi(G, v) \leq \frac{c_9 \, b^2 \log(n)^2 d^{2t} L^{2t+2}}{n}, \quad \|\psi\|_\star \leq \frac{c_{10} \, b^2 \log(n)^3 d^{2t} L^{2t+2}}{n},$$

and as above

$$\sum_{x \in [n]} \mathbb{E}[f(T_x, x)] = \sum_x [P w_x](x) \mu_i^t \mu_j^t \sum_{s=0}^t \frac{[K^s \varphi^{i,j}](x)}{(\mu_i \mu_j)^s}$$

$$= \sum_x [K\mathbf{1}](x) \mu_i^t \mu_j^t \sum_{s=0}^t \frac{[K^s \varphi^{i,j}](x)}{(\mu_i \mu_j)^s} = \Gamma_{V,ij}^{(t)}.$$

Equation (8.10) is then derived as we did earlier.

Our final inequality to prove is (8.11); we consider now the function

$$F_t(g, o) = \mathbf{1}_{(g,o)_{t+1} \text{ has no cycles}} \vec{\partial}_\mathbf{1} [F_{i,t}^2](g, o).$$

For all $t \geq 0$, the function F_t is $t + 1$-local, and

$$|F_t(g, o)| \leq \deg(o) \big(2\|\varphi_i\|_\infty |(g, o)_{t+1}|\big)^2 L^{2t}$$

$$\leq 4 \deg(o) \frac{b^2}{n} |(g, o)_{t+1}|^2 L^{2t} := \psi_t(g, o).$$

Whenever $v \notin \mathcal{V}_t$,

$$F_t(G, v) = \sum_{e:e_2=v} \big([B^{t+1}\chi_i](v) - \mu_i [B^t \chi_i](v)\big)^2.$$

The same computations as in the other equations then imply that

$$\left| \|B^{t+1}\chi_i - \mu_i B^t \chi_i\|^2 - \sum_{x \in [x]} F_t(G, v) \right| \leq 2|\mathcal{V}_t| \max_v \psi(G, v),$$

and

$$\max_v \psi(G, v) \leq \frac{c_5 \, b^2 \log(n)^3 d^{2t+1} L^{2t}}{n}, \quad \|\psi\|_\star \leq \frac{c_5 \, b^2 \log(n)^4 d^{2t+1} L^{2t}}{n}.$$

Furthermore,

$$\sum_{x\in[n]} \mathbb{E}\big[F_t(T_x,x)\big] = \sum_{x\in[n]} [P\mathbf{1}](x)\big[K^{t+1}\varphi^{i,i}\big](x) = \langle P\mathbf{1}, K^{t+1}\varphi^{i,i}\rangle,$$

and we can apply Lemma 3 to find

$$\left|\sum_{x\in[n]} \mathbb{E}\big[F_t(T_x,x)\big]\right| \le rd^3 L^2 \rho^{t+1}.$$

Concluding as above,

$$
\begin{aligned}
\|B^{t+1}\chi_i - \mu_i B^t \chi_i\|^2 &\le rd^3 L^2 \rho^{t+1} + \frac{c_{11}\, b^2 \log(n)^5 d^{3t+3} L^{2t}}{n} \\
&\quad + \frac{c_{12}\, rb^2 \log(n)^7 d^{3t+3} L^{2t}}{\sqrt{n}} \\
&\le rd^3 L^2 \rho^{t+1} + \frac{c_{13} rb^2 d^3 \log(n)^7 d^{3t} L^{2t}}{\sqrt{n}}.
\end{aligned}
$$

D. Proofs for Theorem 10

D.1. Proof of (5.4)–(5.6)

We shall make use of the following classical bound: for a $r_0 \times r_0$ matrix M, we have

$$\|M\| \le r_0 \|M\|_\infty. \tag{D.1}$$

First, the (i,j) entry of matrix U^*U is $\langle u_i, u_j\rangle$, and using (8.9) we find

$$|\langle u_i, u_j\rangle - \Gamma_{U,ij}^{(\ell)}| \le \frac{c\, rb^2 d^2 \log(n)^7 d^{3\ell} L^{2\ell}}{(\mu_i \mu_j)^\ell \sqrt{n}}.$$

Since $i, j \le r_0$, we have $\mu_i \mu_j \ge L^2$, thus

$$|\langle u_i, u_j\rangle - \Gamma_{ij}^{(\ell)}| \le \frac{c\, rb^2 d^2 \log(n)^7 d^{3\ell}}{\sqrt{n}}.$$

By definition of ℓ, it is easy to check that $d^{3\ell} \le n^{-1/4}$. Via (D.1), this implies that $\|U^*U - \Gamma_U^{(\ell)}\|$ is less than $Cn^{-1/4}$, the desired result. The derivation of (5.5) is identical, the bound from Proposition 10 being essentially the same for both cases.

We now move onto the proof of (5.6); we write the scalar product $\langle B^\ell \chi_i, (B^*)^\ell D_W \check{\chi}_j \rangle$ as $\langle B^{2\ell} \chi_i, D_W \check{\chi}_j \rangle$ and use (8.8) to find

$$
\begin{aligned}
|\langle u_i, v_j \rangle - \delta_{ij}| &\le \frac{c\, r b^2 d^3 L \log(n)^6 d^{4\ell} L^{2\ell}}{\mu_i^{2\ell+1} \sqrt{n}} \\
&\le \frac{c\, r b^2 d^{7/2} L \log(n)^6 d^{4\ell}}{\sqrt{n}}.
\end{aligned}
$$

The bound we now need is $d^{4\ell} \le n^{1/4}$, which is true by choice of ℓ, and we conclude as above.

D.2. Bounding $\|B^\ell\|$: proof of (5.9)

Let w be any unit vector in $R^{\vec{E}}$, and assume that we are in the event described in Proposition 3. Then

$$
\begin{aligned}
\|B^t w\|^2 &= \sum_{e \in \vec{E}} \left(\sum_{(e_0,\dots,e_t) \in \mathcal{P}(e,t)} \prod_{i=0}^{t-1} W_{e_i e_{i+1}} w(e_t) \right)^2 \\
&\le L^{2\ell} \sum_{e \in \vec{E}} |\mathcal{P}(e,t)| \sum_{(e_0,\dots,e_t) \in \mathcal{P}(e,t)} w(e_t)^2
\end{aligned}
$$

by the Cauchy–Schwarz inequality. Under the good event from Proposition 3, we have

$$
|\mathcal{P}(e,t)| \le 2|(G,e)_t| \le c_1 \log(n) d^\ell.
$$

Additionally, note that the factor $w(e_t)^2$ appears for each path of length t ending at e_t, or equivalently (reversing edge orientation) for each path in $\mathcal{P}(e_t^{-1}, t)$. Hence,

$$
\begin{aligned}
\|B^t w\|^2 &\le c_1 \log(n) d^\ell L^{2\ell} \sum_{e \in \vec{E}} w(e)^2 |\mathcal{P}(e^{-1}, t)| \\
&\le c_2 \log(n)^2 d^{2\ell} L^{2\ell},
\end{aligned}
$$

and the definition of ℓ ensures (generously) that $d^{2\ell} < \sqrt{n}$.

D.3. Proof of Lemma 9

Note first that for all $t \ge 0$, the parity-time equation (2.2) allows the simplification

$$
\langle (B^*)^t D_W \chi_i, w \rangle = \langle B^t \chi_i, D_W J w \rangle,
$$

and we have $\|D_W J w\| \le L$. Further, the assumption $w \in H^\perp$ implies

$$
\mu_i^{-t} \langle (B^*)^t D_W \chi_i, w \rangle = \mu_i^{-t} \langle (B^*)^t D_W \chi_i, w \rangle - \mu_i^{-\ell} \langle (B^*)^\ell D_W \chi_i, w \rangle;
$$

combining the two above arguments and using a telescopic sum as in the proof of (5.7) gives

$$\left|\mu_i^{-t}\langle (B^*)^t D_W \chi_i, w\rangle\right| = \left|\sum_{s=t}^{\ell-1} \mu_i^{-s}\langle B^s \chi_i, D_W J w\rangle - \mu_i^{-(s+1)}\langle B^{s+1}\chi_i, D_W J w\rangle\right|$$

$$\leq \sum_{s=t}^{\ell-1} \mu_i^{-(s+1)}\left|\langle B^{s+1}\chi_i, D_W J w\rangle - \mu_i\langle B^s\chi_i, D_W J w\rangle\right|$$

$$\leq L \sum_{s=t}^{\ell-1} \mu_i^{-(s+1)}\|B^{s+1}\chi_i - \mu_i B^s \chi_i\|,$$

where we used the Cauchy–Schwarz inequality at the last line. Now, we can apply equation (8.11):

$$\|B^{s+1}\chi_i - \mu_i B^s \chi_i\|^2 \leq rd^3 L^2 \rho^{s+1} + \frac{crb^2 d^3 \log(n)^7 d^{3s} L^{2s}}{\sqrt{n}},$$

and still following the proof of (5.7), we find

$$\|B^{s+1}\chi_i - \mu_i B^s \chi_i\| \leq \sqrt{r}d^{3/2}L\rho^{(s+1)/2} + \frac{c_2 b d^{3/2} \log(n)^{7/2} d^{3s/2} L^s}{n^{1/4}}.$$

Summing these inequalities (and using $\ell \leq c_3 \log(n)$) yields

$$\left|\langle (B^*)^t D_W \check\chi_i, w\rangle\right| \leq \sqrt{r}d^{3/2}L^2 \mu_i^t \sum_{s=t}^{\ell-1}\left(\frac{\sqrt{\rho}}{\mu_i}\right)^{s+1}$$

$$+ \mu_i^{-\ell}\frac{c_4 b d^{3/2} \log(n)^{9/2} d^{3\ell/2} L^\ell}{n^{1/4}}.$$

Since $i \leq r_0$, we have $\mu_i > \sqrt{\rho}$. As a result, all terms in the sum are bounded by the one for $s = t - 1$, and $\mu_i^{-\ell} \leq d^{\ell/2}$. We finally get

$$\left|\langle (B^*)^t D_W \check\chi_i, w\rangle\right| \leq \sqrt{r}d^{3/2}L^2 \rho^{t/2} + \frac{c_4 b d^{3/2} \log(n)^{9/2} d^{2\ell} L^\ell}{n^{1/4}},$$

as desired.

E. Norm bounds: additional proofs

E.1. Bound (9.7) on $\|MB^k\|$

Since $\|M\|$ is of order 1, we notice that (9.7) improves by a factor of $\sqrt{n}$ on the crude bound $\|KB^k\| \leq \|K\|\|B^k\|$. We use the same trace method as above; we have

$$\|MB^{k-2}\|^{2m} \leq \operatorname{tr}\left[\left(MB^{k-2}(B^*)^{k-2}M^*\right)^m\right]$$

$$\leq \left(\frac{d}{n}\right)^{2m} \sum_{\gamma \in W_{k,m}} \prod_{i=1}^{m} X_{\gamma_{2i-1,1}\gamma_{2i-1,2}} \prod_{s=3}^{k} A_{\gamma_{2i-1,s-1}\gamma_{2i-1,s}}$$

$$\times \prod_{s=1}^{k-2} A_{\gamma_{2i,s-1}\gamma_{2i,s}} X_{\gamma_{2i,k-2}\gamma_{2i,k-1}},$$

where $W_{k,m}$ is the set of sequences of paths defined just below equation (10.1). The set of edges of the form $(\gamma_{2i-1,0}, \gamma_{2i-1,1})$ or $(\gamma_{2i,k-1}, \gamma_{2i,k})$, which support no random variable, has cardinality at most m by the boundary conditions, hence the bound for any $\gamma \in W_{k,m}$:

$$\prod_{i=1}^{m} X_{\gamma_{2i-1,1}\gamma_{2i-1,2}} \prod_{s=3}^{k} A_{\gamma_{2i-1,s-1}\gamma_{2i-1,s}} \prod_{s=1}^{k-2} A_{\gamma_{2i,s-1}\gamma_{2i,s}} X_{\gamma_{2i,k-2}\gamma_{2i,k-1}}$$

$$\leq \left(\frac{d}{n}\right)^{e_\gamma - m} L^{2(k-2)m}.$$

Using bound (10.3) on $\mathcal{W}_{k,m}(v, e)$ and the fact that each equivalence class contains at most n^v elements, we get

$$\|MB^{k-2}\|^{2m} \leq \left(\frac{d}{n}\right)^m \sum_{e=1}^{2km} \sum_{v=1}^{e+1} k^{2m}(2km)^{6m(e-v+1)} n^v \left(\frac{d}{n}\right)^e L^{2(k-2)m}$$

$$\leq n^{-m} d^{5m} L^{2km} k^{2m} \sum_{e=1}^{2km} \sum_{v=1}^{e+1} (2km)^{6m(e-v+1)} d^e n^{v-e}$$

$$\leq n^{-m+1} d^{5m} L^{2km} k^{2m} (2km) d^{2km} \sum_{g=0}^{\infty} \left(\frac{(2km)^{6m}}{n}\right)^g. \qquad (E.1)$$

The choice of parameter
$$m = \left\lceil \frac{\log(n)}{12 \log(\log(n))} \right\rceil$$

ensures that the infinite sum in (E.1) converges for n larger than an absolute constant, which yields (9.7).

E.2. Bound (9.8) on $\|\Delta^{(t-1)} \tilde{M} B^{k-t-1}\|$

First, notice that $M_{ef}^{(2)}$ is equal to $(TQT^*)_{ef}$ except when $\mathbf{1}\{e \overset{2}{\to} f\} = 0$, which happens only when $e = f, e \to f, e \to f^{-1}$, or $f^{-1} \to e$. Therefore, we can write

$$|L_{ef}| \leq \frac{d}{n}(\tilde{M}_1 + \tilde{M}_2 + \tilde{M}_3 + \tilde{M}_4),$$

where each entry of the matrix M_i is one whenever the i-th condition mentioned above is true. Then, for each i, we can write

$$\|\Delta^{(t-1)}\tilde{M}_i B^{k-t-1}\| \leq \|\Delta^{(t-1)}\|\|\tilde{M}_i B^{k-t-1}\|,$$

and a straightforward adaptation of the proof of bound (9.7) gives

$$\frac{d}{n}\|\tilde{M}_i B^{k-t-1}\| \leq \frac{cd^{7/2}L\ln(n)^7 d^{k-t}L^{k-t}}{\sqrt{n}}.$$

Combining the above bound with (9.6) easily implies (9.8).

E.3. Bound (9.9) on $R_t^{(\ell)}$

The proof of (9.9) is very similar to those above, as well as the one in [14]; we only highlight the main differences. Let $t \geq 1$ (the case $t = 0$ is almost identical), and $k \leq \log(n)$. The same trace argument gives

$$\|R_t^{(k-1)}\|^{2m} \leq \mathrm{tr}\Big[\big(R_t^{(k-1)}R_t^{(k-1)^*}\big)^m\Big]$$

$$= \sum_{\gamma \in T_{k,m,t}} \prod_{i=1}^{2m} \underline{X}_{\gamma_{i,0}\gamma_{i,1}} \prod_{s=2}^{t} \underline{A}_{\gamma_{i,s-1}\gamma_{i,s}} Q_{\gamma_{i,t},\gamma_{i,t+1}} \prod_{s=t+2}^{k} A_{\gamma_{i,s-1}\gamma_{i,s}},$$

where $T_{k,m,t}$ is the set of sequences of paths $(\gamma_1, \ldots, \gamma_{2m})$ such that for all i, $\gamma_i^1 = (\gamma_{i,0}, \ldots, \gamma_{i,t})$ and $\gamma_i^2 = (\gamma_{i,t+1}, \ldots, \gamma_{i,k})$ are tangle-free and γ_i is tangled, with similar boundary conditions as in (10.2).

   We define G_γ as the union of the $G_{\gamma_i^z}$ for $z \in [2m]$, $j \in \{1, 2\}$. Since we remove an edge to each path, G_γ need not be connected; however, since γ_i is tangled, each connected component in G_{γ_i} contains a cycle, and the same holds for G. It follows that

$$v_\gamma \leq e_\gamma$$

for all $\gamma \in T_{k,m,t}$. As before, we define the equivalence relation $\sim$ and $\mathcal{T}_{k,m,t}(v, e)$ the set of equivalence classes with $v_\gamma = v$ and $e_\gamma = e$. Then, the following lemma from [14] holds:

Lemma 14. *Let v, e be any integers such that $v \leq e$. Then*

$$\big|\mathcal{T}_{k,m,t}(v, e)\big| \leq (4km)^{12m(e-v+1)+8m}.$$

As for bounding the contribution of a single path, the computations already performed in bounding (9.6) work similarly:

$$\prod_{i=1}^{2m} \underline{X}_{\gamma_{i,0}\gamma_{i,1}} \prod_{s=2}^{t} \underline{A}_{\gamma_{i,s-1}\gamma_{i,s}} Q_{\gamma_{i,t},\gamma_{i,t+1}} \prod_{s=t+2}^{k} A_{\gamma_{i,s-1}\gamma_{i,s}}$$
$$\leq \left(\frac{a}{n}\right)^{e_\gamma+2m} \left(1+\frac{d}{n}\right)^{2km} d^{2m} L^{2km},$$

using $Q_{ij} \leq dL/n$ for all i, j. Finally, for $[\gamma] \in \mathcal{T}_{k,m,t}(v, e)$, there are at most n^v sequences γ' such that $\gamma' \sim \gamma$. This yields

$$\mathbb{E}\big[\|R_t^{(k-1)}\|^{2m}\big] \leq c_1^{2m} d^{2m} L^{2km} \left(\frac{d}{n}\right)^{2m} (4km)^{20m} \sum_{e=1}^{2km} (4km)^{12m(e-v)} \sum_{v=1}^{e} d^e n^{v-e}$$
$$\leq c_2^{2m} d^{4m} L^{2km} n^{-2m} \log(n)^{40m} (2km) d^{2km} \sum_{g=0}^{\infty} \left(\frac{(4km)^{12m}}{n}\right)^g,$$

using preemptively the bound $m \leq \log(n)$ and the change of variables $g = e - v$. This time, choosing

$$m = \left\lceil \frac{\log(n)}{24\log(\log(n))} \right\rceil$$

yields a convergent sum, and (9.9) follows.

References

[1] E. Abbe, Community detection and stochastic block models: recent developments. *J. Mach. Learn. Res.* **18** (2017), Paper No. 177   Zbl 1403.62110   MR 3827065

[2] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, Decoding binary node labels from censored edge measurements: phase transition and efficient recovery. *IEEE Trans. Network Sci. Eng.* **1** (2014), no. 1, 10–22   MR 3349181

[3] E. Abbe and C. Sandon, Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *FOCS'15 – IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 670–688, IEEE Computer Soc., Los Alamitos, CA, 2015   MR 3473334

[4] E. Abbe and C. Sandon, Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.* **71** (2018), no. 7, 1334–1406   Zbl 1394.62082   MR 3812075

[5] J. Alt, R. Ducatez, and A. Knowles, Extremal eigenvalues of critical Erdős–Rényi graphs. *Ann. Probab.* **49** (2021), no. 3, 1347–1401   Zbl 1467.05236   MR 4255147

[6] J. Baik, G. Ben Arous, and S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** (2005), no. 5, 1643–1697   Zbl 1086.15022   MR 2165575

[7] H. Bass, The Ihara–Selberg zeta function of a tree lattice. *Internat. J. Math.* **3** (1992), no. 6, 717–797   Zbl 0767.11025   MR 1194071

[8]   F. L. Bauer and C. T. Fike, Norms and exclusion theorems. *Numer. Math.* **2** (1960), 137–141   Zbl 0101.25503   MR 118729

[9]   F. Benaych-Georges, C. Bordenave, and A. Knowles, Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. *Ann. Probab.* **47** (2019), no. 3, 1653–1676   Zbl 1447.60017   MR 3945756

[10] F. Benaych-Georges, C. Bordenave, and A. Knowles, Spectral radii of sparse random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** (2020), no. 3, 2141–2161   Zbl 1459.15036   MR 4116720

[11] F. Benaych-Georges and R. R. Nadakuditi, The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** (2011), no. 1, 494–521   Zbl 1226.15023   MR 2782201

[12] F. Benaych-Georges and R. R. Nadakuditi, The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** (2012), 120–135   Zbl 1252.15039   MR 2944410

[13] C. Bordenave, S. Coste, and R. R. Nadakuditi, Detection thresholds in very sparse matrix completion. 2020, arXiv:2005.06062

[14] C. Bordenave, M. Lelarge, and L. Massoulié, Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs. *Ann. Probab.* **46** (2018), no. 1, 1–71   Zbl 1386.05174   MR 3758726

[15] E. J. Candes and Y. Plan, Matrix completion with noise. *Proc. IEEE* **98** (2010), no. 6, 925–936

[16] M. Capitaine, C. Donati-Martin, and D. Féral, The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** (2009), no. 1, 1–47   Zbl 1163.15026   MR 2489158

[17] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** (2013), Paper No. 59   Zbl 1373.15053   MR 3068390

[18] U. Feige and E. Ofek, Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** (2005), no. 2, 251–275   Zbl 1076.05073   MR 2155709

[19] D. Féral and S. Péché, The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** (2007), no. 1, 185–228   Zbl 1136.82016   MR 2291807

[20] Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrices. *Combinatorica* **1** (1981), no. 3, 233–241   Zbl 0494.15010   MR 637828

[21] S. Heimlicher, M. Lelarge, and L. Massoulié, Community detection in the labelled stochastic block model. 2012, arXiv:1209.2910

[22] P. W. Holland, K. B. Laskey, and S. Leinhardt, Stochastic blockmodels: First steps. *Social Networks* **5** (1983), no. 2, 109–137   MR 718088

[23] I. M. Johnstone and D. Paul, PCA in high dimensions: An Orientation. *Proc. IEEE* **106** (2018), no. 8, 1277–1292

[24] R. H. Keshavan, A. Montanari, and S. Oh, Matrix completion from noisy entries. In *NeurIPS 2009 – Advances in Neural Information Processing Systems 22*, pp. 952–960, Curran Associates, Inc., 2009

[25] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA* **110** (2013), no. 52, 20935–20940 Zbl 1359.62252 MR 3174850

[26] J. Lei and A. Rinaldo, Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** (2015), no. 1, 215–237 Zbl 1308.62041 MR 3285605

[27] M. Lelarge, L. Massoulié, and J. Xu, Reconstruction in the labelled stochastic block model. *IEEE Trans. Network Sci. Eng.* **2** (2015), no. 4, 152–163 MR 3453283

[28] L. Massoulié, Community detection thresholds and the weak Ramanujan property. In *STOC'14 – Proceedings of the 2014 ACM Symposium on Theory of Computing*, pp. 694–703, ACM, New York, 2014 Zbl 1315.68210 MR 3238997

[29] A. Montanari and S. Sen, Semidefinite programs on sparse random graphs and their application to community detection. In *STOC'16 – Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 814–827, ACM, New York, 2016 Zbl 1376.90043 MR 3536616

[30] S. Péché, The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** (2006), no. 1, 127–173 Zbl 1088.15025 MR 2221787

[31] A. Saade, M. Lelarge, F. Krzakala, and L. Zdeborová, Spectral detection in the censored block model. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1184–1188, IEEE, 2015

[32] L. Stephan and L. Massoulié, Robustness of spectral methods for community detection. In *COLT 2019 – 32nd Annual Conference on Learning Theory*, pp. 2831–2860, Proceedings of Machine Learning Research 99, PMLR, 2019

[33] Y. Watanabe and K. Fukumizu, Graph zeta function in the Bethe free energy and Loopy belief propagation. In *NeurIPS 2009 – Advances in Neural Information Processing Systems 22*, pp. 2017–2025, Curran Associates, Inc., 2009

[34] H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.* **71** (1912), no. 4, 441–479 Zbl 43.0436.01 MR 1511670

[35] Y. Yu, T. Wang, and R. J. Samworth, A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102** (2015), no. 2, 315–323 Zbl 1452.15010 MR 3371006

Ludovic Stephan

Information, Learning and Physics (IdePHICS) lab, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland; ludovic.stephan@epfl.ch

Laurent Massoulié

Inria Paris – Département d'informatique de l'ENS, PSL Research University, 75012 Paris, France; laurent.massoulie@inria.fr