Annals of Mathematics **190** (2019), 811–875
https://doi.org/10.4007/annals.2019.190.3.3

Eigenvalues of random lifts and polynomials of random permutation matrices

By Charles Bordenave and Benoît Collins

Abstract

Let $(\sigma_1, \ldots, \sigma_d)$ be a finite sequence of independent random permutations, chosen uniformly either among all permutations or among all matchings on n points. We show that, in probability, as $n \to \infty$, these permutations viewed as operators on the $n-1$ dimensional vector space $\{(x_1, \ldots, x_n) \in \mathbb{C}^n, \sum x_i = 0\}$, are asymptotically strongly free. Our proof relies on the development of a matrix version of the non-backtracking operator theory and a refined trace method.

As a byproduct, we show that the non-trivial eigenvalues of random n-lifts of a fixed based graphs approximately achieve the Alon-Boppana bound with high probability in the large n limit. This result generalizes Friedman's Theorem stating that with high probability, the Schreier graph generated by a finite number of independent random permutations is close to Ramanujan.

Finally, we extend our results to tensor products of random permutation matrices. This extension is especially relevant in the context of quantum expanders.

1. Introduction

1.1. *Weighted sum of permutations.* Let X be a countable set and r, d positive integers. We consider $(a_0, \ldots, a_d)$ elements in $M_r(\mathbb{C})$ and let $\sigma_i \in S(X), i \in \{1, \ldots, d\}$, be permutations of the set X. Let $\ell^2(X)$ be the Hilbert space spanned by an orthonormal basis δ_x indexed by the elements of $x \in X$. The permutation σ_i acts naturally as a unitary operator S_i on $\ell^2(X)$ by σ_i, $S_i(g)(x) = g(\sigma_i(x))$ for all $g \in \ell^2(X)$. Let 1 be the identity operator on $\ell^2(X)$. We are interested in the operator on $\mathbb{C}^r \otimes \ell^2(X)$,

$$A = a_0 \otimes 1 + \sum_{i=1}^{d} a_i \otimes S_i. \tag{1}$$

If X is a finite set with n elements, A may be viewed as an $rn \times rn$ matrix.

Keywords: random permutation, strong asymptotic freeness, random lifts, non-backtracking operator

AMS Classification: Primary: 60B20, 46L54, 05C80.

If X is finite, the constant vector $\mathbf{1} \in X$ is in $\ell^2(X)$. In addition, it is an eigenvector of any permutation matrix of X associated to the eigenvalue 1. Therefore,

$$H_1 = \mathbb{C}^r \otimes \mathbf{1}$$

is an invariant vector space of A and A^* of dimension r. The restriction of A to H_1 is given by

$$A_1 = a_0 + \sum_{i=1}^{d} a_i. \tag{2}$$

When X is finite, we are interested in the spectrum of A on $H_0 = H_1^\perp$. The space H_0 is the set of $g \in \mathbb{C}^r \otimes \ell^2(X)$ such that $\sum_x g(x) = 0 \in \mathbb{C}^r$, where $g(x)$ denotes the orthogonal projection of g on $\mathbb{C}^r \otimes \{\delta_x\}$ — canonically identified to $\mathbb{C}^r$. We note that A leaves H_0 invariant, i.e., $AH_0 \subset H_0$, therefore it also defines an operator on H_0. We will denote by $A_{|H_0}$ the restriction of A to H_0, which we see as an element of $B(H_0)$.

We use the following standard notation. For a positive integer n, wet set $[n] = \{1, \dots, n\}$. The *absolute value* of a bounded operator T is $|T| = \sqrt{TT^*}$, $\sigma(T)$ denotes the *spectrum* of T and $\|T\|$ its *operator norm*. Finally, $\rho(T)$ is the *spectral radius* of T, and if T is self-adjoint, we set $s(T) = \sup \sigma(T)$ (the right edge of the spectrum). For example,

$$\|A_{|H_0}\| = \sup_{g : g \in H_0 \setminus \{0\}} \frac{\|Ag\|}{\|g\|}.$$

It will be often useful to work with the operator A when it is self-adjoint. Self-adjointness is ensured by replacing the operators S_i by generic algebraically free unitary operators and checking self-adjointness. This condition is essentially sufficient. (It is sufficient if the cardinal of X is large enough and if one requires the property to hold for any choice of permutation S_i.) For practical purposes, let us assume that the set $\{1, \dots, d\}$ is endowed with an involution $i \mapsto i^*$ (and $i^{**} = i$ for all i). Then, the symmetry condition is fulfilled as soon as

$$a_0^* = a_0 \quad \text{and} \quad \text{for all } i \in \{1, \dots, d\}, \quad (a_i)^* = a_{i^*} \quad \text{and} \quad \sigma_{i^*} = \sigma_i^{-1}. \tag{3}$$

If the symmetry condition (3) holds, then A is self-adjoint. In this case,

$$s(A_{|H_0}) = \sup_{g : g \in H_0 \setminus \{0\}} \frac{\langle g, Ag \rangle}{\|g\|^2}.$$

We are interested in the spectrum of A when $X = [n]$, the permutations σ_i are random and n is large. The operator A becomes a random matrix that we study in the case when σ_i, $i \in [d]$ are random permutations whose distribution are described below.

Definition 1 (Symmetric random permutations). For some integer $0 \le q \le d/2$, we have for $1 \le i \le q$, $i^* = i + q$, for $q+1 \le i \le 2q$, $i^* = i - q$ and for $2q+1 \le i \le d$, $i^* = i$. The permutations $(\sigma_i), i \in \{1, \dots, q\} \cup \{2q+1, \dots, d\}$, are independent, for $1 \le i \le 2q$, σ_i is uniformly distributed in $\mathcal{S}_n$ and $\sigma_{i^*} = \sigma_i^{-1}$. If $2q < d$, then we assume that n is even and for $2q+1 \le i \le d$, σ_i is a uniformly distributed matching on $[n]$ (where a matching is a permutation σ such that $\sigma^2(x) = x$ and $\sigma(x) \ne x$ for all $x \in [n]$).

1.2. *Large n limit, non-commutative probability spaces.* The following operator describes the local limit of A (in the sense of Benjamini-Schramm; see [3], [8]) as n grows large. For symmetric random permutations, let q be as in Definition 1 and $X = X_\star = \mathbb{Z} * \cdots * \mathbb{Z} * \mathbb{Z}_2 * \cdots * \mathbb{Z}_2$ be the free product generated by q copies of $\mathbb{Z}$ and $d - 2q$ copies of $\mathbb{Z}_2$. We denote by $g_1, \dots, g_d$ its generators, where if $1 \le i \le q$, then (g_i, g_{i+q}) generates the i-th copy of $\mathbb{Z}$. In $\mathbb{C}^r \otimes \ell^2(X_\star)$, we define the convolution operator

$$A_\star = a_0 \otimes 1 + \sum_{i=1}^{d} a_i \otimes \lambda(g_i), \tag{4}$$

with $a_0, \dots, a_d$ from equation (1), and where $\lambda(g)$ is the left regular representation (left multiplication).

In the non-commutative probability vocabulary, $A_\star$ is called a *non-commutative random variable*; namely, it is an element of $\mathcal{A}$, where $\mathcal{A}$ is a unital *-algebra and τ is a faithful trace on it. Here, $\mathcal{A}$ is $M_r(C_r(X_\star))$, where $C_r(X_\star)$ is the reduced C^*-algebra of the group $X_\star$ and the trace is $r^{-1}\mathrm{Tr} \otimes \tau$, where $\tau(\lambda(g)) = \mathbf{1}(g = e)$.

Recall that a sequence of complex random variables (Y_n) converges *in probability* to $y \in \mathbb{C}$ if for any $\varepsilon > 0$, $\mathbb{P}(|Y_n - y| \ge \varepsilon)$ converges to 0 as n goes to infinity.

1.3. *Linear or not linear*? In this paper we study the spectrum of the operator A defined in equation (1). The spectrum of the limiting operator defined in equation (4) gives a candidate for the limiting spectrum, which we will show to be correct with high probability. This operator A is a *linear* combination of the permutation matrices S_i's with matrix coefficients. On the other hand, the abstract of this manuscript mentions strong asymptotic freeness. As defined below, this is a property that involves the behavior of any non-commutative *polynomial* in the variables S_i's with scalar coefficient; i.e., it is *not* necessarily a linear combination of the S_i's. So there is no obvious a priori implication between the two problems. It turns out that these questions are actually equivalent. This fact is an important phenomenon that has been widely used in random matrix theory in the last two decades, known as the *linearization trick*. Details are provided in Section 6.

1.4. *Large n limit, main result.* For symmetric random permutations, it follows from results of Nica [34] that the operators $(S_i), i \in [d]$, are *asymptotically free* in probability. This means that for any polynomial P in unitaries $\lambda(g_i)$ $((g_i), i \in [d]$, symmetric generators of the group $X_\star$, as per the definition above equation (4), the corresponding polynomial P_n obtained by replacing $\lambda(g_i)$ by S_i (seen as a random variable permutation on $M_n(\mathbb{C})$) satisfies that the random variable $n^{-1}\mathrm{Tr}(P_n) \to \tau(P)$, where this convergence holds in probability. This notion is a particular case of the concept of *asymptotic freeness*. A good and modern introduction can be found in [32]. Although the results of this paper can also be seen as a contribution to the asymptotic theory of freeness, a non-expert reader can safely assume that the explanations developed in this paragraph cover the necessary background in free probability.

This notion of convergence proved by Nica for the permutations operators S_i as the dimension n grows to infinity shows that for any self-adjoint polynomial P_n in S_i, the percentage of eigenvalues in a given real interval $[a, b]$ converges to the spectral measure of the limiting polynomial P on the group $X_\star$ on the same interval $[a, b]$. In particular, if $[a, b]$ does not intersect the limiting spectrum, it shows that the percentage of eigenvalues in this interval tends to zero. But it does not rule out the possibility for $o(n)$ eigenvalues being in this interval. If such eigenvalues exist, they are called *outliers*. As a matter of fact, in our model, outliers can be made to exist by taking an appropriate polynomial and the constant vector $\mathbf{1}$. For example, $S_1 + S_1^* + \cdots + S_k + S_k^*$ always has an eigenvalue $2k$, and this is an outlier as soon as $k \geq 2$.

It is very natural to ask whether there are more outliers than those potential obvious ones. For some random matrix models, it was shown that this is not the case. For example, a (negative) answer to the unitary version of this problem was achieved by the second author and Male [14], as a continuation of the seminal result of Haagerup and Thorbjørnsen [24]. The proofs are based on the linearization trick that reduces such a question to the analog question on polynomials of degree one and with matrix coefficients, as our operator A in (1).

Whenever there are no outliers in a limit of a (random) matrix model, one says that it converges *strongly*. Mathematically, it is equivalent to saying that the norm of any polynomial converges to the supremum of its limiting spectrum. Specifically, beyond assuming the existence of a limit of $n^{-1}\mathrm{Tr}(P_n)$ for any polynomial, one assumes in addition that

$$\lim_n ||P_n|| = \lim_\ell (\lim_n n^{-1}\mathrm{Tr}((P_n P_n^*)^\ell)^{(2\ell)^{-1}}. \tag{5}$$

(Note that this notion is not probabilistic — when the operators P_n are random, e.g., because they are constructed out of random unitaries, then one may consider such notions of convergences to a constant in some probabilistic sense, for example, in probability, or almost surely.)

The above ideas are well captured by stating that the spectrum of a self-adjoint operator is not far from its limiting spectrum in the sense of the Hausdorff distance. Recall that the *Hausdorff distance* between two subsets S and T of $\mathbb{R}$ is the infimum over all $\varepsilon > 0$ such that $S \subset T+[-\varepsilon,\varepsilon]$ and $T \subset S+[-\varepsilon,\varepsilon]$. Let us also remark that it is not completely obvious at first sight that the notion introduced in equation (5) (i.e., convergence of the operator norm) and the notion of convergence of spectrum in Hausdorff distance are equivalent. One has to check the absence of outliers *between* two connected components of the limiting spectrum if it is not connected. This happens to be equivalent because the quantifier for strong convergence runs over *every* polynomial. We refer to Section 6 for additional details.

THEOREM 2. *If the symmetry condition* (3) *holds, for symmetric random permutations, as n goes to infinity, the Hausdorff distance between $\sigma(A_{|H_0})$ and $\sigma(A_\star)$ converges in probability to* 0. *In particular, $s(A_{|H_0})$ converges in probability to $s(A_\star)$ in the sense of the Hausdorff distance.*

We note that there is an explicit expression for $\|A_\star\|$ and $s(A_\star)$ in the self-adjoint case. The scalar case $r = 1$ is due to Akemann and Ostrand [2], and the general case for any r and a_i Hermitian is due to Lehner [28].

A corollary of this result is

THEOREM 3. *For symmetric random permutations, the permutation operators restricted to $\mathbf{1}^\perp$, $((S_i)_{|\mathbf{1}^\perp}), i \in [d]$, are asymptotically strongly free in probability.*

1.5. *Spectral gaps of random graphs.* In equation (1), consider the special case where $a_0 = 0$ and for any $i \in [d]$, $a_i = E_{u_i v_i}$ for some u_i, v_i in $[r]$ (where $(E_{uv})_{u'v'} = \mathbf{1}_{(u,v)=(u',v')}$). Then A is the adjacency matrix of a colored graph on the vertex set $[n] \times [r]$ and whose directed edges with color $i \in [d]$ are $((x, u_i), (\sigma_i(x), v_i))$, for all $x \in [n]$. If the symmetry condition (3) holds, then this graph is undirected. This graph is called a *n-lift of the base graph* whose adjacency matrix $A_1 = \sum_i a_i$ is given by (2); see Figure 1 for a concrete example.

For random symmetric permutations, the n-lift is random. Since the work of Amit and Linial [4], [5] and Friedman [17], this class of random graphs has attracted a substantial attention [29], [1], [30], [38], [35], [20]. The Alon-Boppana lower bounds asserts that for any $\varepsilon > 0$, for all n large enough and any permutations $(\sigma_i), i \in [d]$, in $\mathcal{S}_n$, with the symmetry condition (3), we have

$$s(A_{|H_0}) \geq s(A_\star) - \varepsilon \tag{6}$$

(due in this context to Greenberg [22]). Then, Theorem 2 proves that random n-lifts achieve the Alon-Boppana lower bound (6) up to a vanishing term. It

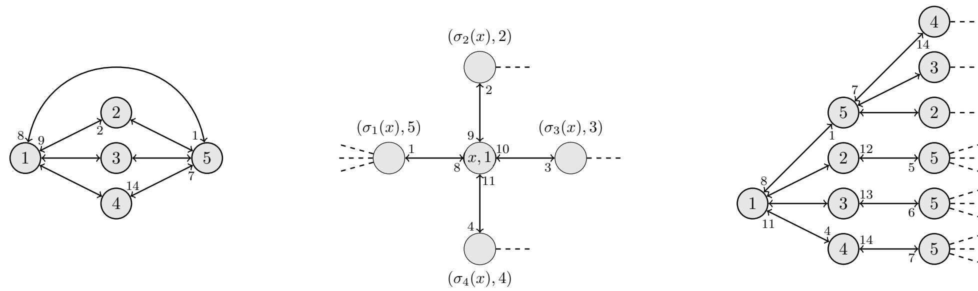


Figure 1. Left: An undirected base graph with $r = 5$ vertices and $q = d/2 = 7$ edges, with q as in Definition 1. We have $a_1 = E_{15}$, $a_2 = E_{12}$, $a_3 = E_{13}$, $a_4 = E_{14}$, $a_5 = E_{25}$, $a_6 = E_{35}$, $a_7 = E_{45}$ and for $8 \leq i \leq 14$, $a_i = a^*_{i-7} = a^*_{i^*}$. The subscripts on the arrows are the index of the corresponding $i \in [d]$. (They are not all represented for the sake of readability.) Middle: Neighborhood in the n-lift of a vertex $(x, 1)$. Right: Picture of the common universal covering tree of the base graph and the n-lifts.

follows that A has up to vanishing terms the largest possible spectral gap (the difference between the largest eigenvalue and the second largest). It settles the conjecture in Friedman [17], and it proves an even stronger statement, i.e., *all* eigenvalues of $A_{|H_0}$ are ε-close to the spectrum of $A_\star$; see Figure 2 for a numerical illustration. In some cases, it was already proved in Friedman [18] ($r = 1$, $a_i = 1$), Friedman and Kohler [20] and Bordenave [9] ($r \geq 1$, $a_i = E_{u_i v_i}$, $A_1 = \sum_i a_i$ constant row sum) and, up to a multiplicative factor, in Puder [38].

Now, consider the case, $r = 1$, $a_i \geq 0$ and $\sum_i a_i = 1$.T then A is the Markov transition matrix of an *anisotropic random walk*. For the properties of this random walk on the free group, see the monograph by Figà-Talamanca and Steger [16]. More generally, for any integer $r \geq 1$, if $A_1 = \sum_i a_i$ is a stochastic matrix, then A is also a stochastic matrix that can be interpreted as a Markov chain on the n-lift graph. The Alon-Boppana lower bounds (6) holds also in this context; see Ceccherini-Silberstein, Scarabotti and Tolli [12]. Thus, Theorem 2 asserts that A has, up to vanishing terms, the largest possible spectral gap. Interestingly, our argument will actually show that (10) is achieved with probability tending to one, jointly for all a_i with $\max \|a_i\| \leq 1$.

In the same vein, assume that all a_i, $i \in [d]$, have non-negative entries and $a_0 = -\sum_i a_i$. Then A is a Laplacian matrix and it is the infinitesimal generator of a continuous time random walk on the n-lift. Theorem 2 proves again that, up to vanishing terms, random permutations maximize the spectral gap of such processes.

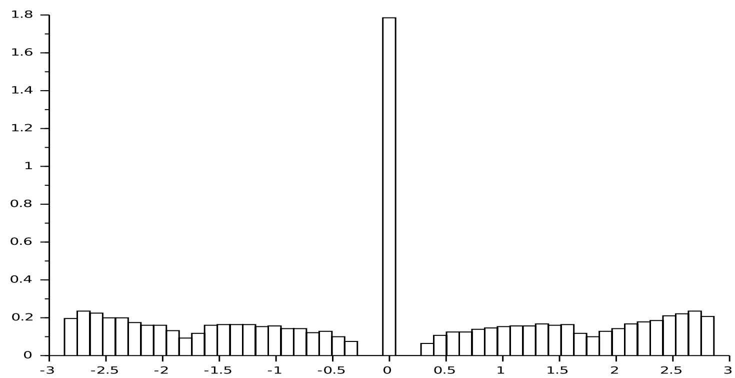


Figure 2. *Histogram of the eigenvalues of* $A_{|H_0}$ *for* $n = 500$ *and a random* n*-lift of the base graph depicted in Figure* 1: The spectrum of $A_\star$ is the spectrum of the universal covering tree of the base graph: we have $\sigma(A_\star) = [-a, -b] \cup \{0\} \cup [b, a]$ with $a \simeq 2.866$ and $b \simeq 0.283$.

1.6. *Tensor product of random permutation matrices.* We now discuss an extension of our work that is notably relevant in the context of quantum expanders; see [25], [26] and in cryptography [19]. Let X be a finite set and r, d positive integers. Let $\ell^2(X^2)$ be the Hilbert space spanned by an orthonormal basis $\delta_{(x,y)}$ indexed by the elements of $(x, y) \in X^2 = X \times X$. We consider $(a_0, \ldots, a_d)$ elements in $M_r(\mathbb{C})$ and let $\sigma_i \in S(X)$ be a permutation of the set X whose corresponding unitary operator on $\ell^2(X)$ is S_i. We are now interested in the operator on $\mathbb{C}^r \otimes \ell^2(X^2)$,

$$A^{(2)} = a_0 \otimes 1 \otimes 1 + \sum_{i=1}^{d} a_i \otimes S_i \otimes S_i. \tag{7}$$

Note that (7) is again an operator of the form (1) since $S_i \otimes S_i$ is a unitary operator associated to the permutation on $\ell^2(X^2)$ defined for all $(x, y) \in X^2$ by $\sigma_i^{(2)}(x, y) = (\sigma_i(x), \sigma_i(y))$. Note also that we may identify $\ell^2(X^2)$ with linear operators on $\ell^2(X)$ endowed with the Hilbert-Schmidt scalar product $\langle a, b\rangle = \mathrm{Tr}(a^* b)$. Then $S_i \otimes S_i$ is identified with the linear map $T \mapsto S_i T S_i^*$.

We consider the following orthogonal elements of $\ell^2(X^2)$, defined in coordinates, for all $(x, y) \in X^2$,

$$J_{xy} = \mathbf{1}(x \neq y) \quad \text{and} \quad I_{xy} = \mathbf{1}(x = y). \tag{8}$$

It is immediate to check that for any permutation operator S on $\ell^2(X)$,

$$(S \otimes S)(J) = J \quad \text{and} \quad (S \otimes S)(I) = I.$$

Let V be the vector space spanned by I and J. We introduce the vector space of codimension $2r$,

$$H_0^{(2)} = \mathbb{C}^r \otimes V^\perp.$$

For random permutations, we have the following analog to Theorem 2.

THEOREM 4. *Theorem 2 holds with A replaced by $A^{(2)}$ defined by (7) and H_0 replaced by $H_0^{(2)}$.*

In Section 5, we will explain how to adapt the proof of Theorem 2 to deal with tensor products. Interestingly, the analog of Theorem 4 for random unitaries is not known and it cannot be deduced from [14], [24]. As corollary of Theorem 4, we have the following

THEOREM 5. *For symmetric random permutations, the permutation operators restricted to $V^\perp$, $((S_i \otimes S_i)_{|V^\perp}), i \in [d]$, are asymptotically strongly free in probability.*

Note that asymptotic freeness follows, among others, from [13]. In particular, when one restricts oneself to sum of generators $\sum_i S_i \otimes S_i$, one obtains that the family $S_i \otimes S_i$ viewed as an operator on $V^\perp$ is a nearly optimal quantum expander in the sense of Hastings [25] and Pisier [37].

1.7. *Brief overview.* The proof of Theorem 2 is divided into two parts. For any $\varepsilon > 0$, we will prove that with probability tending to one,

(1) the spectrum of $A_\star$ is contained in an arbitrarily small neighborhood of the spectrum of $A_{|H_0}$,

$$\sigma(A_\star) \subset \sigma(A_{|H_0}) + [-\varepsilon, \varepsilon]; \tag{9}$$

and

(2) the spectrum of $A_{|H_0}$ is contained in an arbitrarily small neighborhood of the spectrum of $A_\star$,

$$\sigma(A_{|H_0}) \subset \sigma(A_\star) + [-\varepsilon, \varepsilon]. \tag{10}$$

The proof of the spectrum inclusion (9) is standard and follows from the asymptotic freeness of independent permutations, which follows from the already mentioned reference, Nica [34]. However we give an alternative argument by supplying a general deterministic criterion that guarantees that (9) holds and we prove that the symmetric random permutations meet this criterion.

The proof of (10) is much more involved, and it is the main contribution of this work. It relies on a novel use of *non-backtracking operators.* These operators are defined on an enlarged vector space and, despite the fact that they are non-normal, they are much easier to work with. Indeed, they have a very simple form on the free product of groups $X_\star$. Notably, in Theorem 12, we prove that the set $\sigma(A)\backslash\sigma(A_\star)$ is controlled by the spectral radii of a one-parameter family of non-backtracking operators. This will allow us to reduce the proof of Theorem 2 to the proof of Theorem 17, which is an analogous statement for non-backtracking operators. Then, the proof of Theorem 17 follows a strategy similar to the new proof of Friedman's Theorem in [9] but with non-negligible refinements.

Indeed, the main technical novelty will be the presence of matrix-valued weights a_i, $i \in [d]$. In particular, we are able to relate directly the expectation of the trace of a power of a non-backtracking matrix on $[n]$ to powers of the corresponding non-backtracking operator on the free group $X_\star$ (forthcoming Lemma 26). Another important issue will be that we will need a refined net argument to control jointly the spectral radii of the non-backtracking matrices associated to all possible weights a_i with uniformly bounded norms. Due to the non-normality of non-backtracking matrices, we have to deal with the bad regularity of spectral radii in terms of matrix entries. See Remark 1 for a more precise comparison with previous works.

The remainder of this manuscript is organized as follows. In Section 2, we prove the spectrum inclusion (9). In Section 3, we introduce non-backtracking operators and we reduce the proof of Theorem 2 to Theorem 17 on non-backtracking operators. In Section 4, we prove this last theorem on non-backtracking operators for random permutation matrices. In Section 5, we adapt the previous arguments to prove Theorem 4. The proof of Theorems 3 and 5 is contained in Section 6. It is based on the linearization trick. It was developed simultaneously in various areas of mathematics, and applied in operator algebras by Haagerup and Thorbjørnsen, and subsequently it was improved by Anderson [7]. For a synthetic introduction we will refer to Mingo and Speicher [32, p. 256]. Finally, proofs of auxiliary results are gathered in Section 7.

Notation. We use the usual notation $o(\cdot)$ and $O(\cdot)$. We denote by $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ the corresponding probability measures on $\mathcal{S}_n^d$, corresponding to the definition of equation (1). Note that $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ depend implicitly on n. The coordinate vector at $x \in X$ is denoted δ_x. It will be convenient to describe our operators as matrix-valued operators. For an operator M on $\mathbb{C}^r \otimes \ell^2(X)$, we set for all $x, y \in X$,

$$M_{xy} = (1 \otimes \langle \delta_x, \cdot \delta_y \rangle)(M) \in M_r(\mathbb{C}). \tag{11}$$

In other words, we may see M as an infinite block matrix indexed by $X \times X$ (of matrices in $M_r(\mathbb{C})$), and we may reformulate the above equation as $M = (M_{xy})_{(x,y)\in X\times X}$. Finally, we will use the convention that a product over an empty set is equal to 1 and the sum over an empty set is 0.

Acknowledgements. CB was supported by ANR-16-CE40-0024-01 and ANR-14-CE25-0014. He would like to thank Doron Puder for early discussions on this problem. BC was supported by JSPS KAKENHI 17K18734, 17H04823, 15KK0162 and ANR-14-CE25-0003. He would like to thank Mikael de la Salle, Camille Male and Amir Dembo for enriching discussions on random permutations and free probability. Both authors are grateful to Gilles Pisier for discussions during the finalization phase of the paper, including useful references and questions that encouraged us to write Theorem 5. The final version of this paper was completed during the visit of CB to Kyoto University under a JSPS short term professorship. Both authors gratefully acknowledge the support of JSPS and Kyoto University. The authors are indebted to Yu Shang-Chun for his careful reading. Finally, the authors would like to thank an anonymous referee for an extremely detailed and insightful report that allowed them to improve greatly their initial manuscript.

2. Inclusion of the spectrum of $A_\star$

Assume that $X = [n]$ and that the symmetry condition (3) holds. We start by some elementary definitions from graph theory and define a natural colored graph G^σ associated to the permutations σ_i, $i \in [d]$.

Definition 6.

- A *colored edge* $[x, i, y]$ is an equivalence class of triplets $(x, i, y) \in [n] \times [d] \times [n]$ endowed with the equivalence $(x, i, y) \sim (x', i', y')$ if $(x', i', y') \in \{(x, i, y), (y, i^*, x)\}$. A *colored graph* H is a graph whose vertices is a subset of $[n]$ and whose edges are a set of colored edges.
- A *path* in H of length k from x to y is a sequence $(x_1, i_1, \dots, x_{k+1})$ such that $x_1 = x$, $x_{k+1} = y$ and $[x_t, i_t, x_{t+1}]$ is an edge of H for any $1 \le t \le k$. The path is *closed* if $x_1 = x_{k+1}$. A *cyclic path* in H is a closed path in H such that $(x_1, \dots, x_k)$ are pairwise distinct. A *cycle* is the colored graph spanned by a cyclic path. (That is, the vertex set is $\{x_t : 1 \le t \le k\}$ and the edge set is $\{[x_t, i_t, x_{t+1}] : 1 \le t \le k\}$.)
- If $x \in [n]$ and h is a non-negative integer, then $(H, x)_h$ denotes the subgraph of H spanned by all edges belonging to a path starting from x and of length at most h.
- G^σ is the colored graph whose vertex set is $[n]$ and whose edges are the set of $[x, i, y]$ such that $\sigma_i(x) = y$ (and $\sigma_{i^*}(y) = x$).

The inclusion (9) is a direct consequence of the following deterministic proposition whose proof can be found in Section 7 for completeness.

PROPOSITION 7. *Assume that $X = [n]$ and that the symmetry condition (3) holds. Let A and $A_\star$ be defined by (1) and (4). For any $\varepsilon > 0$, there exists* *an integer* $h \geq 1$ *such that if* $(G^\sigma, x)_h$ *contains no cycle for some* $x \in [n]$, *then* $\sigma(A_\star) \subset \sigma(A_{|H_0}) + [-\varepsilon, \varepsilon]$.

As a corollary, we obtain the first half of Theorem 2.

COROLLARY 8. *Let* A *and* $A_\star$ *be as in* (1) *with the symmetry condition* (3). *For symmetric random permutations, for any* $\varepsilon > 0$, *with probability tending to one as* n *goes to infinity*, $\sigma(A_\star) \subset \sigma(A_{|H_0}) + [-\varepsilon, \varepsilon]$.

Proof. If $x \in [n]$ is such that $(G^\sigma, x)_h$ contains a cycle, we claim that there exists a cycle of length $k \leq 2h$ contained in $(G^\sigma, x)_h$. Indeed, assume that $(G^\sigma, x)_h$ contains a cycle, and let y be a vertex on this cycle that is at maximal distance $t \leq h$ from x (where the distance is the minimal t such that there exists a path of length t from x to y). Fix a path γ from x to y of length t. At least one of two neighboring edge of y on the cycle is not on γ; we call it $[y, i, y']$. Let γ' be a path from y' to x of minimal length $t' \leq t$. By construction, $[y, i, y']$ is not an edge of γ'. It follows that the sequence (γ, i, γ') forms a closed path in $(G^\sigma, x)_h$ that contains a cycle. It is of length at most $t + t' + 1 \leq 2h$. (If $t = h$, then $t' < t$ by the definition of $(G^\sigma, x)_h$.)

Now, for an integer $k \geq 1$, let N_k be the number of distinct cycles of length k in G^σ. In the forthcoming Lemma 23, we will check that the expectation of N_k is $O((d-1)^k)$. Also, for a given cycle in G^σ of length $k \leq 2h$, there are at most $d(d-1)^{h-1}$ vertices $x \in [n]$ such that $(G^\sigma, x)_h$ contains this cycle. It follows that the expected number of $x \in [n]$ such that $(G^\sigma, x)_h$ contains a cycle is upper bounded by

$$\sum_{k=1}^{2h} d(d-1)^{h-1} \mathbb{E} N_k = O((d-1)^{3h}).$$

(This bound is very rough.) Hence, by Markov inequality, for any $h \leq (\log n)/\kappa$ with $\kappa > 3 \log(d-1)$, there exists, with probability tending to one, a vertex $x \in [n]$ such that $(G^\sigma, x)_h$ contains no cycle. On the latter event, we may conclude by applying Proposition 7. □

3. Non-backtracking operator

3.1. *Spectral mapping formulas.* Let A be as in (1). We consider a vector space U of finite codimension in $\ell^2(X)$ that is left invariant by all permutation operators S_i, $i \in [d]$: $S_i U = U$. We set

$$H = \mathbb{C}^r \otimes U.$$

This vector space H is left invariant by A.

We define $E = X \times [d]$. If A is thought of as a weighted adjacency operator of a directed graph on the vertex set X, an element (x, i) of E can be thought as a directed edge from x to $\sigma_i(x)$ with weight a_i (where all vertices have a loop edge of weight a_0). The non-backtracking operator B associated to A is the operator on $\mathbb{C}^r \otimes \ell^2(E) = \mathbb{C}^r \otimes \ell^2(X) \otimes \mathbb{C}^d$ defined by

$$B = \sum_{j \neq i^*} a_j \otimes S_i \otimes E_{ij}, \tag{12}$$

where $E_{ij} \in M_d(\mathbb{R})$ is the matrix defined by $(E_{ij})_{kl} = \mathbf{1}_{(i,j)=(k,l)}$ for all $j, k \in [d]$. Equivalently, writing B as a matrix-valued operator on $\ell^2(E)$, we have for any $e, f \in E$ and $e = (x, i) \in E$, $f = (y, j)$,

$$B_{ef} = a_j \mathbf{1}(\sigma_i(x) = y)\mathbf{1}(j \neq i^*).$$

See Figure 3 for an informal illustration of the operator A and its non-backtracking operator B.

Note that B does not depend on the matrix element a_0. We set

$$K = \mathbb{C}^r \otimes U \otimes \mathbb{C}^d.$$

We observe from (12) that B defines an operator on K, that is, $BK \subset K$. As before, we denote by $B_{|K}$ the restriction of B to K.

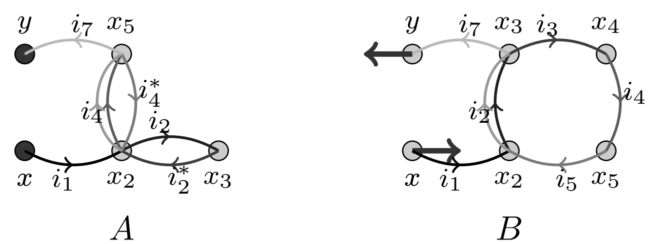


Figure 3. Left: For $x, y \in X$ and k integer, $(A^k)_{xy} \in M_r(\mathbb{C})$ is the sum of all weighted paths $(x_1, i_1, x_2, \ldots, i_k, x_{k+1})$ in G^σ (as per Definition 6 with extra loop edges of weight a_0 at all vertices) from $x = x_1$ to $y = x_{k+1}$ with weight $\prod_{t=1}^k a_{i_t}$. In the example, we have $k = 7$ and the path is such that $i_3 = i_2^*$, $i_5 = i_4^*$, $i_6 = i_4$. Right: For $e = (x, i), f = (y, j) \in E = X \times [d]$ and k integer, $(B^k)_{ef} \in M_r(\mathbb{C})$ is the sum of all weighted paths $(x_1, i_1, x_2, \ldots, i_k, x_{k+1})$ in G^σ with $(x_1, i_1) = e$, $(x_{k+1}, i_{k+1}) = f$ where $i_{k+1} = j$, and for all $1 \leq t \leq k$, $i_{t+1} \neq i_t^*$. The weight of the path is $\prod_{t=1}^k a_{i_{t+1}}$. In the example, we have $k = 7$ and $i_6 = i_2$. The condition $i_{t+1} \neq i_t^*$ is viewed as a non-backtracking constraint of the path. If $X = X_\star$ with generators $(g_i)_{i \in [d]}$ as defined above (4) and $\sigma_i = \lambda(g_i)$, then the condition $i_{t+1} \neq i_t^*$ asserts that $g_{i_1} \cdots g_{i_{k+1}}$ is in reduced form.

The following statement relates the spectrum of B with the spectrum of an operator of the same type as A. In the scalar case $r = 1$, the next proposition is contained in Watanabe and Fukumizu [39].

PROPOSITION 9. *Let A be as in (1) with associated non-backtracking operator B, and let $\lambda \in \mathbb{C}$ satisfy $\lambda^2 \notin \{\sigma(a_i a_{i^*}) : i \in [d]\}$. Define the operator A_λ on $\mathbb{C}^r \otimes \ell^2(X)$ through*

$$
\begin{aligned}
A_\lambda &= a_0(\lambda) + \sum_{i=1}^{d} a_i(\lambda) \otimes S_i, \\
a_i(\lambda) &= \lambda a_i (\lambda^2 - a_{i^*} a_i)^{-1}, \\
a_0(\lambda) &= -1 - \sum_{i=1}^{d} a_i (\lambda^2 - a_{i^*} a_i)^{-1} a_{i^*}.
\end{aligned} \tag{13}
$$

Then $\lambda \in \sigma(B)$ if and only if $0 \in \sigma(A_\lambda)$. Moreover, $\lambda \in \sigma(B_{|K})$ if and only if $0 \in \sigma((A_\lambda)_{|H})$.

Proof of Proposition 9. We first assume λ is in the discrete spectrum of B. We show that 0 is in the discrete spectrum of A_λ. Then, there is an eigenvector $v \in \mathbb{C}^r \otimes \ell^2(E)$ such that $Bv = \lambda v$, which reads in the coordinates of $\ell^2(E)$, for all $e = (x, i) \in E$,

$$
\lambda v(x, i) = \sum_{j \neq i^*} a_j v(\sigma_i(x), j), \tag{14}
$$

with $v(e) \in \mathbb{C}^r$. We define $u \in \mathbb{C}^r \otimes \ell^2(X)$ by, for each $x \in X$,

$$
u(x) = \sum_j a_j v(x, j)\,. \tag{15}
$$

(If elements (x, i) of E are thought as derivatives at x in a discrete direction i, then the vector u can be thought as a divergence vector). The eigenvalue equation (14) reads

$$
\lambda v(x, i) = u(\sigma_i(x)) - a_{i^*} v(\sigma_i(x), i^*)\,.
$$

Applying the above identity to $e = (\sigma_i(x), i^*) = (\sigma_{i_*}^{-1}(x), i^*)$, we find

$$
\lambda v(\sigma_i(x), i^*) = u(x) - a_i v(x, i)\,.
$$

We deduce

$$
\lambda^2 v(x, i) = \lambda^2 u(\sigma_i(x)) - a_{i^*} u(x) + a_{i^*} a_i v(x, i)\,.
$$

By assumption, $\lambda^2 - a_{i^*} a_i$ is invertible. Hence,

$$
v(x, i) = (\lambda^2 - a_{i^*} a_i)^{-1} (\lambda u(\sigma_i(x)) - a_{i^*} u(x))\,. \tag{16}
$$

Let us note that equations (15) and (16), when restricted, define a map between H_0 and K_0. We see from this last expression that $u \neq 0$ if $v \neq 0$. We now check that u is in the kernel of A_λ. Let $y \in X$, $i \in [d]$, and set $x = \sigma_i^{-1}(y) = \sigma_{i^*}(y)$. We plug (16) into (14) and get

$$\begin{aligned}\lambda^2(\lambda^2 - a_{i^*}a_i)^{-1}u(y) - \lambda(\lambda^2 - a_{i^*}a_i)^{-1}a_{i^*}u(x) \\ = \sum_{j\neq i^*} \lambda a_j(\lambda^2 - a_{j^*}a_j)^{-1}u(\sigma_j(y)) - \sum_{j\neq i^*} a_j(\lambda^2 - a_{j^*}a_j)^{-1}a_{j^*}u(y).\end{aligned}$$

Since $\sigma_{i^*}(y) = x$, we find

$$\begin{aligned}\lambda^2(\lambda^2 - a_{i^*}a_i)^{-1}u(y) = \sum_{j} \lambda a_j(\lambda^2 - a_{j^*}a_j)^{-1}u(\sigma_j(y)) \\ - \sum_{j\neq i^*} a_j(\lambda^2 - a_{j^*}a_j)^{-1}a_{j^*}u(y).\end{aligned}$$

Since $1 = \lambda^2\big(\lambda^2 - a_{i^*}a_i\big)^{-1} - a_{i^*}\big(\lambda^2 - a_i a_{i^*}\big)^{-1}a_i$, we conclude that

$$\left(1 + \sum_j a_j(\lambda^2 - a_{j^*}a_j)^{-1}a_{j^*}\right)u(y) = \sum_j \lambda a_j(\lambda^2 - a_{j^*}a_j)^{-1}u(\sigma_j(y)),$$

which proves that u is the kernel of A_λ.

Conversely, if 0 is in the discrete spectrum of A_λ with eigenvector u, we define v through (16). (Note that $v \neq 0$ because of (15).) Then the above computation also implies that v satisfies (14), i.e., $Bv = \lambda v$, so that λ is in the discrete spectrum of B. Note also that $u \in H_0$ if and only if $v \in K_0$.

Finally, if λ is in the essential spectrum of B, then, for any $\varepsilon > 0$, there exists $v \in \mathbb{C}^r \times \ell^2(E)$ such that $\|v\|_2 = 1$ and $\|Bv - \lambda v\|_2 \leq \varepsilon$. The above argument shows that $\|A_\lambda u\|_2 = O(\varepsilon)$ and (16) implies that $\|v\|_2 = O(\|u\|_2)$. It implies that 0 is in the spectrum of A_λ. Conversely, if 0 is in the essential spectrum of A_λ, then λ is in the spectrum of B. □

This proposition could be used for studying the spectrum of the operator A. To this end, we should for a given A and μ find a corresponding B_μ such that μ is the spectrum of A if and only if $\lambda = 1$ is in the spectrum of B_μ. Assume that there are q pairs $\{i, i^*\}$ such that $i \neq i^*$ and p elements of $[d]$ such that $i = i^*$, with $2q + p = d$. For concreteness, as in Definition 1, we may assume without loss of generality that for $1 \leq i \leq q$, $i^* = i + q$, for $q + 1 \leq i \leq 2q$, $i^* = i - q$ and for $2q + 1 \leq i \leq d$, $i^* = i$. Now, let $A_\star$ be as in (4). We relate the spectra of A and B through the resolvent of $A_\star$. More precisely, for $\mu \notin \sigma(A_\star)$, we set

$$G(\mu) = (\mu - A_\star)^{-1}.$$

In the symmetric and scalar case $r = 1$, the next proposition is a formula derived in Anantharaman [6, §7].

For the next proposition and for the sequel, we recall that we use a matrix notation with indices in $X \times X$ for operators on $\mathbb{C}^r \otimes \ell^2(X)$, as per equation (11).

In particular, $G(\mu)$ defined above fits in this context with $X = X_\star$, and it will be written as $G_{xy}(\mu) = (G(\mu))_{xy}$. We will be interested in the case where x and y are o (the neutral element of $X_\star$ for its group structure) or g_i (the i-th generator of $X_\star$ defined above (4)).

Let D be a bounded set in $\mathbb{C}$. We define $\mathrm{full}(D) = \mathbb{C}\backslash U$, where U is the unique infinite component of $\mathbb{C}\backslash D$. (Informally stated, $\mathrm{full}(D)$ fills the holes of D.) For example, if $D \subset \mathbb{R}$ or D is simply connected, then $\mathrm{full}(D) = D$.

PROPOSITION 10. *Let A be as in (1) and $\mu \notin \mathrm{full}(\sigma(A_\star))$. Define the operator A_μ on $\mathbb{C}^r \otimes \ell^2(X)$ through*

$$A_\mu = \sum_{i=1}^{d} \hat{a}_i(\mu) \otimes S_i, \qquad \text{with} \quad \hat{a}_i(\mu) = G_{oo}(\mu)^{-1} G_{og_i}(\mu). \tag{17}$$

Let B_μ be the corresponding non-backtracking operator. Then $\mu \notin \sigma(A)$ if and only if $1 \notin \sigma(B_\mu)$. Moreover, $\mu \notin \sigma(A_{|H})$ if and only if $1 \notin \sigma((B_\mu)_{|K})$.

We start with a classical expression for $\hat{a}_i(\mu)$ related to the recursive equations satisfied by resolvent operators on trees.

LEMMA 11. *Let $A_\star$ be as above, and let $\mu \notin \mathrm{full}(\sigma(A_\star))$ and $\hat{a}_i(\mu) = G_{oo}(\mu)^{-1} G_{og_i}(\mu)$. Then the following identities hold*:

$$G_{oo} = \left(\mu I_r - a_0 - \sum_i \hat{a}_i a_{i^*} \right)^{-1}$$

and

$$a_i G_{oo} = \hat{a}_i (I_r - \hat{a}_{i^*} \hat{a}_i)^{-1}.$$

Proof. By analyticity, it is sufficient to prove the identities for all μ in a neighborhood of infinity. We introduce the operator $A_\star^{(o)}$ on $\mathbb{C}^r \otimes \ell^2(X)$ defined by

$$A_\star - A_\star^{(o)} = \sum_j a_j \otimes \delta_o \otimes \delta_{g_j} + a_{j^*} \otimes \delta_{g_j} \otimes \delta_o.$$

In words, $A_\star^{(o)}$ is the operator associated to the Cayley graph of $X_\star$ where the the unit o has been isolated. We denote by $G^{(o)}$ the resolvent of $A_\star^{(o)}$. If μ is large enough, we have $\mu \notin \sigma(A_\star^{(o)})$. We set $\gamma_i(\mu) = G^{(o)}_{g_i g_i}(\mu)$. Then, omitting μ for ease of notation, the resolvent identity reads

$$G = G^{(o)} + G(A_\star - A_\star^{(o)}) G^{(o)} = G^{(o)} + G^{(o)}(A_\star - A_\star^{(o)}) G.$$

Observe that $G^{(o)}_{og_i} = 0$ and $G^{(o)}_{g_j g_i} = 0$ for $j \neq i$ (since there is a direct decomposition of $A_\star^{(o)}$ on $\ell^2(X_\star) = \mathbb{C}\delta_o \oplus_i \ell^2(g_i X_\star)$). Thus, composing the resolvent identity by $(1 \otimes \langle \delta_o, \cdot\, \delta_{g_i} \rangle)$ and $(1 \otimes \langle \delta_{g_i}, \cdot\, \delta_o \rangle)$, we obtain $G_{og_i} = G_{oo} a_i \gamma_i$

and $G_{g_i o} = \gamma_i a_{i^*} G_{oo}$. Then applying the last inequality to $g_{i^*} = g_i^{-1}$ and using that $G_{xg,yg} = G_{xy}$ for any x, y, g, it gives the formula

$$G_{og_i} = G_{oo} a_i \gamma_i = \gamma_{i^*} a_i G_{oo}.$$

We may now prove the first formula of the lemma. We use that $G^{(o)}_{oo} = (\mu - a_0)^{-1}$ and now compose the resolvent identity by $(1 \otimes \langle \delta_o, \cdot \delta_o \rangle)$. We obtain

$$G_{oo} = G^{(o)}_{oo} + \sum_j G_{og_j} a_{j^*} G^{(o)}_{oo} = (\mu - a_0)^{-1} + \sum_j G_{oo} a_j \gamma_j a_{j^*} (\mu - a_0)^{-1}.$$

Multiplying on the right by $\mu - a_0$, we obtain

$$G_{oo} \left(\mu - a_0 - \sum_j a_j \gamma_j a_{j^*} \right) = 1,$$

as claimed.

For the second formula, we first repeat the above argument with

$$A_\star^{(og_i)} = A_\star^{(o)} - \sum_{j \neq i^*} a_j \otimes \delta_{g_i} \otimes \delta_{g_j g_i} - a_{j^*} \otimes \delta_{g_j g_i} \otimes \delta_{g_i}.$$

We use the resolvent identity between $A_\star^{(o)}$ and $A_\star^{(og_i)}$. Using that $G^{(og_i)}_{g_j g_i, g_j g_i} = \gamma_j$, we find that$\gamma_i = \left(\mu - a_0 - \sum_{j \neq i^*} a_j \gamma_j a_{j^*} \right)^{-1}$. It implies that

$$\gamma_i (1 - a_{i^*} \gamma_{i^*} a_i \gamma_i)^{-1} = (\gamma_i^{-1} - a_{i^*} \gamma_{i^*} a_i)^{-1} = \left(\mu - a_0 - \sum_j a_j \gamma_j a_{j^*} \right)^{-1} = G_{oo},$$

which concludes the proof. □

Proof of Proposition 10. Consider the operator $B = B_\mu$, and let A_1 be as in Proposition 9. For all $i \in [d]$, we have $a_i(1) = \hat{a}_i (1 - \hat{a}_{i^*} \hat{a}_i)^{-1}$. It is sufficient to prove that $A_1 = (A - \mu)(G_{oo} \otimes 1)$. Using Lemma 11, we deduce that $a_i(1) = a_i G_{oo}$. In particular,

$$a_0(1) = -1 - \sum_i \hat{a}_i (1 - \hat{a}_{i^*} \hat{a}_i)^{-1} \hat{a}_{i^*} = -1 - \sum_i a_i G_{oo} G^{-1}_{oo} G_{og_{i^*}}.$$

Then using Lemma 11 again, we find

$$a_0(1) = -1 - \sum_i a_i G_{og_{i^*}}. = - \left(G^{-1}_{oo} + \sum_i \hat{a}_i a_{i^*} \right) G_{oo} = (a_0 - \mu) G_{oo},$$

as desired. □

3.2. *Spectral radius of non-backtracking operators.* The next theorem is very important in our argument. It gives a sharp criterion to guarantee in terms of non-backtracking operators that the spectrum of an operator A is in a neighborhood of the spectrum of the operator $A_\star$.

THEOREM 12. *The following two results hold true*:

(i) *Let A be as in* (1) *and $A_\star$ the corresponding free operator defined by* (4). *For any $\mu \notin \mathrm{full}(\sigma(A_\star))$, we have $\rho((B_\star)_\mu) < 1$, where $B_\star$ is defined as B in equation (12) with S_i replaced by $\lambda(g_i)$.*

(ii) *For any $\varepsilon > 0$, there exists $\delta > 0$ such that if for all complex μ,*

$$\rho(B_\mu) < \rho((B_\star)_\mu) + \delta,$$

then $\mathrm{full}(\sigma(A))$ is in an ε-neighborhood of $\mathrm{full}(\sigma(A_\star))$.

Moreover, the same holds with $A_{|H}$ and $(B_\mu)_{|K}$.

LEMMA 13. *Let $A_\star$ be as in* (4) *and $B_\star$ its corresponding non-backtracking operator. We have*

$$\{z \in \mathbb{C} : |z| = \rho(B_\star)\} \subset \sigma(B_\star).$$

Proof. If $\rho(B_\star) = 0$, there is nothing to prove. We may thus assume

$$\rho(B_\star) = 1.$$

We start by a consequence of Gelfand's formula on the spectral radius. Let $k \geq 1$ be an integer, and let M be a bounded operator on $\mathbb{C}^k \otimes \ell^2(X_\star)$ in the $C^\star$-algebra generated by operators of the form $b \otimes \lambda(g)$. We introduce the standard tracial state τ defined by

$$\tau(M) = \Big(\frac{1}{k}\mathrm{Tr}(\cdot) \otimes \langle \delta_o, \cdot\, \delta_o\rangle\Big)(M). \tag{18}$$

Gelfand's formula asserts that $\rho(M) = \lim_n \|M^n\|^{1/n}$; see, for example, [33, Th. 1.3.6]. Moreover, since $\|M\|^2 \geq \tau(|M|^2)$, we find

$$\limsup_n \tau(|M^n|^2)^{\frac{1}{2n}} \leq \rho(M).$$

On the other hand, Haagerup's inequality (for matrix valued operators, see Buchholz [11]) asserts that

$$\|M\|^2 \leq c(m) \sum_{x \in X_\star} \|M_{ox}\|^2 = c(m)\|\tau(|M|^2)\| \leq kc(m)\tau(|M|^2),$$

where $m = \sup\{|x| : M_{ox} \neq 0\}$ and $c(m)$ grows polynomially with m (and depends implicitly on k). Hence,

$$\rho(M)^n \leq \|M^n\|^2 \leq kc(nm)\tau(|M^n|^2).$$

Since $c(nm)^{1/n}$ converges to 1 as n grows to infinity, we find

$$\rho(M) \leq \liminf_n \tau(|M^n|^2)^{\frac{1}{2n}}.$$

So finally

$$\rho(M) = \lim_n \tau(|M^n|^2)^{\frac{1}{2n}}.$$

As a consequence, writing $B_\star$ as a convolution operator on $\mathbb{C}^r \otimes \ell^2(X_\star) \otimes \mathbb{C}^d$, we get that

$$1 = \rho(B_\star) = \lim_n \tau(|B_\star^n|^2)^{\frac{1}{2n}}. \tag{19}$$

In particular, there exists $c > 0$ such that for any integer $n \geq 0$,

$$\tau(|B_\star^n|^2) \geq c(1+\varepsilon)^{-2n}. \tag{20}$$

Now, we observe that for any $f \in \mathbb{C}^r$,

$$B_\star^n(f \otimes \delta_{(o,i)}) = \sum_x \left(\prod_{s=2}^{n+1} a_{x_s} f \right) \otimes \delta_x, \tag{21}$$

where the sum is over all reduced words in $X_\star$ $x = (x_1, \ldots, x_{n+1})$ of length $n+1$ with $x_1 = g_i$ and where we have set $a_{g_j} = a_j$ for $j \in [d]$. The lemma is a consequence of the fact that the vectors $B_\star^n(f \otimes \delta_{(o,i)})$, $n \geq 0$, are orthogonal. More precisely, we find for $n \neq m$,

$$\tau(B_\star^n (B_\star^m)^*) = 0. \tag{22}$$

Now, let $z \in \mathbb{C}$ with $|z| = 1 + \varepsilon > \rho(B_\star) = 1$ and $R = (z - B_\star)^{-1}$ be the resolvent of $B_\star$. It is sufficient to check that

$$\tau(|R(z)|^2) \geq c_1 \varepsilon^{-1}$$

for some $c_1 > 0$. (Indeed it implies that $\|R(z)\|^2 \geq \tau(|R(z)|^2)$ diverges as $|z|$ gets closer to unit disc.) Since $|z| > \rho(B_\star)$, we have the converging Taylor expansion

$$R(z) = \sum_n z^{-n-1} B_\star^n.$$

From (20)–(22), we find

$$\tau(|R(z)|^2) = \sum_n |z|^{-2n-2} \tau(|B_\star^n|^2) \geq c \sum_n (1+\varepsilon)^{-4n-2} \geq c_1 \varepsilon^{-1}$$

for some $c_1 > 0$. It gives the desired bound. □

LEMMA 14. *The map $(a_1, \ldots, a_d) \mapsto \rho(B_\star)$ is continuous for any norm on $M_r(\mathbb{C})^d$.*

Before proving Lemma 14, which requires some preliminaries, let us prove Theorem 12.

Proof of Theorem of 12. Let us first prove (ii) assuming (i). By Lemma 14, there exists $\delta > 0$ such that $\rho((B_\star)_\mu) < 1 - \delta$ for all μ at distance larger than ε from $\mathrm{full}(\sigma(A_\star))$. Hence for all μ at distance larger than ε from $\mathrm{full}(\sigma(A_\star))$, we have $\rho(B_\mu) < 1$ and thus by Proposition 10, $\mu \notin \sigma(A)$.

As for (i), assume on the contrary that there exists $\mu_0 \notin \mathrm{full}(\sigma(A_\star))$ such that $\rho(B_{\mu_0}) \geq 1$. Since $\mu_0 \notin \mathrm{full}(\sigma(A_\star))$, there exists a continuous function $t \mapsto \mu_t$, from $[0, \infty)$ to $\mathbb{C} \backslash \mathrm{full}(\sigma(A_\star))$ such that $|\mu_t|$ goes to infinity as t goes to infinity. Note that also $\rho((B_\star)_\mu) \leq \|(B_\star)_\mu\|$ goes to 0 with $|\mu|$ going to infinity. Therefore, Lemma 14 and the intermediate value theorem imply that there exists $t \geq 0$ such that $\mu_t \notin \mathrm{full}(\sigma(A_\star))$ and $\rho((B_\star)_{\mu_t}) = 1$. Then, by Lemma 13, $1 \in \sigma((B_\star)_{\mu_t})$ and, by Proposition 10, we deduce that $\mu_t \in \sigma(A_\star)$. This is a contradiction since $\sigma(A_\star) \subseteq \mathrm{full}(\sigma(A_\star))$. □

Let us now prove Lemma 14. We start with a preliminary statement that is a non-commutative finite-dimensional Perron-Frobenius Theorem due to Krein and Rutman [27]. Let $m \geq 1$ be an integer, and let L be a linear operator on $M_m(\mathbb{C})$. We endow $M_m(\mathbb{C})$ with the standard inner product

$$\langle x, y \rangle = \mathrm{tr}(x^* y).$$

The Frobenius norm is the associated hilbertian norm: $\|x\|_2 = \sqrt{\mathrm{tr}(x^* x)}$. We say that L is of non-negative type if for any x positive semi-definite, Lx is also positive semi-definite. We start with an elementary property of non-negative operators.

LEMMA 15. *Assume that L is of non-negative type. Then L maps hermitian matrices to hermitian matrices. Moreover, for any integer $n \geq 0$, L^n and L^* are of non-negative types.*

Proof. Since L maps positive semi-definite matrices to positive semi-definite matrices, it also maps negative semi-definite matrices to negative semi-definite matrices. Consequently, writing an hermitian matrix as $x = a - b$ with a, b positive semi-definite, we find L maps hermitian matrices to hermitian matrices.

By induction on n, it is immediate from the definition that L^n is also of non-negative type. For L^*, let us first check that it maps hermitian matrices to hermitian matrices. First, from what precedes, $(Ly)^* = L(y^*)$. (We write $y = y_1 + iy_2$ with y_1, y_2 hermitian and use linearity and Ly_i hermitian.) Note that a matrix x is hermitian if and only if for any matrix y, $\mathrm{tr}(x^* y) = \mathrm{tr}(xy)$. Since $\mathrm{tr}(xy) = \mathrm{tr}(yx)$, we get that x is hermitian if and only if $\langle x, y \rangle = \langle y^*, x \rangle$. Let x be hermitian. For any y, we have $\langle L^* x, y \rangle = \langle x, Ly \rangle = \langle (Ly)^*, x \rangle = \langle L(y^*), x \rangle = \langle y^*, L^* x \rangle$, where we have used at the second identity that for any y, $\langle x, y \rangle = \langle y^*, x^* \rangle$ and $x = x^*$. We thus have proved that L^* maps hermitian matrices to hermitian matrices.

Similarly, an hermitian matrix x is positive semi-definite if and only if for any y positive semi-definite, $\langle x, y\rangle = \mathrm{tr}(xy) = \mathrm{tr}(y^{1/2}xy^{1/2}) \geq 0$. However, if x, y are positive semi-definite, then $\langle L^* x, y\rangle = \langle x, Ly\rangle \geq 0$, since L is non-negative. This concludes the proof. □

The following theorem is a direct consequence of the Krein-Rutman Theorem [27]; see, e.g., Deimling [15, Th. 19.1]. For completeness, we have included a proof in Section 7.

THEOREM 16. *Assume that L is of non-negative type, and let ρ be its spectral radius.*

(i) *ρ is an eigenvalue of L and it has a positive semi-definite eigenvector.*
(ii) *If x is positive definite, we have*

$$\lim_{n\to\infty} \|L^n x\|_2^{\frac{1}{n}} = \rho.$$

We are ready for the proof of Lemma 14.

Proof of Lemma 14. From (19), we have $\rho(B_\star) = \lim_{n\to\infty} \|\tau_i(|B_\star^n|^2)\|^{1/2n}$ for some $i \in [d]$. With the notation of (21), we find

$$\tau_i(|B_\star^n|^2) = \tau_i(B_\star^n (B_\star^n)^*) = \sum_x \left| \prod_{s=2}^{n+1} a_{x_s} \right|^2,$$

where the sum is over all reduced words $x = (x_1, \ldots, x_{n+1})$ in $X_\star$ of length $n+1$ with $x_1 = g_i$. Let $m = dr$, and let Z_n be the block diagonal matrix in $M_m(\mathbb{C})$ with diagonal blocks in $M_r(\mathbb{C})$, $(\tau_1(|B_\star^n|^2), \ldots, \tau_d(|B_\star^n|^2))$. We find $Z_0 = 1$ and for integer $n \geq 0$,

$$Z_{n+1} = L(Z_n), \tag{23}$$

where L is the operator on $M_m(\mathbb{C})$ defined as follows. For $x \in M_m(\mathbb{C})$, we write in $x = (x_{ij})$, $i, j \in [d]$ for its blocks in $M_r(\mathbb{C})$. Then $L(x)$ is block diagonal with diagonal blocks $(L(x)_{11}, \ldots, L(x)_{dd})$, and for all i in $[d]$,

$$L(x)_{ii} = \sum_{j \neq i^*} a_j x_{jj} a_j^* \in M_r(\mathbb{C}). \tag{24}$$

It is straightforward to check that L is of non-negative type. Indeed, if x is positive semi-definite then, for each $j \in [d]$, x_{jj} is also positive semi-definite in $M_r(\mathbb{C})$ and thus $a_j x_{jj} a_j^*$ is positive semi-definite.

Now, from (23), $Z_n = L^n(1)$. We deduce from Theorem 16(ii) that $\rho(B_\star)$ is equal to the spectral radius of L. We recall finally that the spectral radius is a continuous function on $M_m(\mathbb{C})$ for any norm. □

3.3. *Random weighted permutations.* We now consider symmetric random permutations, $X = [n]$. We consider the vector space K_0 of vectors $f \in \mathbb{C}^r \otimes \ell^2(E)$ such that $\sum_x f(x,i) = 0$ for all $i \in [d]$ (that is, the orthogonal of $\mathbb{C}^r \otimes \mathbf{1} \otimes \mathbb{C}^d$). The following result is the central technical contribution hidden behind the proof of Theorem 2.

THEOREM 17. *For any $0 < \varepsilon < 1$, for symmetric random permutations, with probability tending to one as n goes to infinity, for all $(a_i), i \in [d]$, such that $\max_i(\|a_i\| \vee \|a_i^{-1}\|^{-1}) \le \varepsilon^{-1}$ and that satisfy the symmetry condition (3), we have*

$$\rho(B_{|K_0}) \le \rho(B_\star) + \varepsilon,$$

where B is the non-backtracking operator associated to A defined by (1) and $\rho(B_\star)$ is the spectral radius of the corresponding non-backtracking operator in the free group.

Note that in the above theorem, the assumption $\max_i(\|a_i\| \vee \|a_i^{-1}\|^{-1}) \le \varepsilon^{-1}$ entails a control on the norm of a_i^{-1} and, in particular, the assumption that it is invertible. This is a technical assumption that does appear in the main result Theorem 2. This will, however, not be a major obstacle for proving Theorem 2 in the next subsection by using the fact that invertible matrices are dense in the space of all matrices $M_r(\mathbb{C})$.

As a corollary, in the next subsection we obtain a proof of our Theorem 2. The forthcoming Section 4 is devoted to the proof of Theorem 17. It relies on a refinement of the trace method Füredi and Komlós [21] that was developed in [31], [10], [9]. In special case where a_i is $E_{x_i y_i}$, this theorem is contained in [9] and under an extra assumption in [20].

3.4. *Proof of Theorem 2.*

Proof of Theorem 2. From Corollary 8, it remains to prove the upper bound (10). The proof is a combination of Theorems 17 and 12 applied to $H = H_0$ and $K = K_0$. We may assume that $\max_i \|a_i\| \le 1$. If A and A' are operators of the form (1) with associated matrices (a_i) and (a'_i) and the same (S_i), we have $\|A - A'\| \le \sum_i \|a'_i - a_i\|$. Hence, up to modifying ε in $\varepsilon/2$, in order to prove the upper bound (10), it is enough to consider weights (a_i) such that for any i, $\|a_i\| \le 1$ and $\|a_i^{-1}\| \le 2d/\varepsilon$ and check that on an event of high probability the upper bound (10) holds. We already know that $\|A_{|H_0}\| \le d$. Let μ be a complex number. Recall that $\hat{a}_i(\mu) = a_i \gamma_i(\mu)$. Moreover, if μ is at distance at least ε from $\sigma(A_\star)$, we have $(|\mu| + d)^{-1} I_r \le |\gamma_i(\mu)| \le \varepsilon^{-1} I_r$. Hence, if $|\mu| \le d$, we get $\|\hat{a}_i(\mu)\| \le \|a_i\|\|\gamma_i(\mu)\| \le 1/\varepsilon$ and $\|\hat{a}_i(\mu)^{-1}\| \le \|a_i^{-1}\|\|\gamma_i(\mu)^{-1}\| \le (2d)^2/\varepsilon$. It remains to use the event of high probability in Theorem 17 with $\varepsilon' = (\varepsilon/(2d)^2) \wedge \delta$, where $\delta > 0$ is as in Theorem 12. □

4. Proof of Theorem 17

4.1. *Overview of the proof.* Let us first describe the method introduced by Füredi and Komlós [21] to bound the norm of a random matrix. Let M be a random matrix in $M_n(\mathbb{C})$. Imagine that we want to prove that, for some $\rho > 0$, for any $\varepsilon > 0$, with probability tending to 1 as n goes large,

$$\|M\| \leq \rho(1+\varepsilon). \tag{25}$$

For integer $k \geq 1$, we write

$$\|M\|^{2k} = \|MM^*\|^k = \|(MM^*)^k\| \leq \operatorname{tr}\Big((MM^*)^k\Big).$$

At the last step, we might typically loose a factor proportional to n, since the trace is the sum of n eigenvalues. Hence, it is reasonable to target a bound of the form

$$\mathbb{E}\operatorname{tr}\Big((MM^*)^k\Big) \leq n\rho^{2k}(1+\varepsilon)^{2k}. \tag{26}$$

If we manage to establish such an upper bound, we would deduce from the Markov inequality that for any $\delta > 0$, the event

$$\|M\| \leq \rho(1+\varepsilon)(1+\delta)$$

has probability at least

$$1 - n(1+\delta)^{-2k} = 1 - \exp\left(-2k\log(1+\delta) + \log n\right).$$

Hence, the bound on the trace (26) implies the bound (25) with $\varepsilon' = \varepsilon + o(1)$ if $k \gg \log n$. Then, the problem of bounding the norm of M has been reduced to bounding the expression

$$\mathbb{E}\operatorname{tr}\Big((MM^*)^k\Big) = \sum_{x_1,\dots,x_{2k}} \mathbb{E}\prod_{t=1}^{k} M_{x_{2t-1},x_{2t}}\bar{M}_{x_{2t+1},x_{2t}},$$

where the sum is over all sequences $(x_1, \dots, x_{2k})$ in $[n]$ with the convention $x_{2k+1} = x_1$. The right-hand side of the above expression may usually be studied by combinatorial arguments.

As it is described, the method of Füredi and Komlós cannot be applied directly in this context. Indeed, assume for concreteness that $r = 1$, that A_1 and A are stochastic matrices, and that the symmetry condition (3) holds. (Recall that $A_1 = \sum_i a_i$ is defined in (2).) We are then interested in the matrix $M = A - \frac{1}{n}\mathbf{1}\otimes\mathbf{1}$ and $\|A_{|H_0}\| = \|M\|$, and we aim at a bound of the type (25) for some $\rho < 1$. However, with probability at least $c_0 n^{-c_1}$, $\{1,2\}$ is a connected component of G^σ, where G^σ is the colored graph introduced in Definition 6. (This can be checked from the forthcoming computation leading to (35).) On this event, say E, the eigenvalue 1 has multiplicity at least 2 in A, hence $\|M\| = 1$. We deduce that

$$\mathbb{E}\operatorname{tr}\Big((MM^*)^k\Big) \geq \mathbb{E}\|M\|^{2k} \geq \mathbb{P}(E) \geq c_0 n^{-c_1}.$$

However, the latter is much larger than ρ^{2k} if $k \gg \log n$ and (26) does not hold. More generally the presence of subgraphs with many edges in G^σ prevents the bound (26) to hold for $k \gg \log n$. Such subgraphs were called *tangles* by Friedman [18]. We will circumvent this intrinsic difficulty as follows. Let B be as in Theorem 17.

(i) *Bound the spectral radius by the norm of a large power*: We fix a positive integer ℓ. We recall that K_0 is the vector space of codimension rd orthogonal to $K_1 = \mathbb{C}^r \otimes \mathbf{1} \otimes \mathbb{C}^d$. We have

$$\rho(B_{|K_0}) = \rho((B^\ell)_{|K_0})^{1/\ell} \leq \sup_{g \in K_0, \|g\|_2 = 1} \|B^\ell g\|_2^{1/\ell}. \tag{27}$$

(ii) *Remove the tangles* (*Section* 4.2): We obtain a matrix $B^{(\ell)}$ that coincides with B^ℓ on an event of high probability.

(iii) *Project on K_0* (*Section* 4.2): We project $B^{(\ell)}$ on K_0 to evaluate the right-hand side of (27). We obtain a matrix $\underline{B}^{(\ell)}$. However, since K_0 is not always an invariant subspace of $B^{(\ell)}$, there will also be some remainder matrices.

(iv) *Method of Füredi and Komlós* (*Section* 4.4): We may then evaluate the norm of $\underline{B}^{(\ell)}$ by taking a trace of power $2m$ and estimate its expectation. The $2m\ell$ plays the role of $2k$ in the above presentation of the method of Füredi and Komlós. We thus need $2m\ell \gg \log n$ to get a sharp estimate. We obtain $2m\ell$ of order $(\log n)^2/(\log \log n)$. This is for the part of the proof where we have previously removed the tangles (in the proofs of Lemmas 25 and 29). We also connect the expected trace of powers of non-backtracking matrices to powers of the corresponding non-backtracking operator on the free group (Lemma 27).

(v) *Net argument* (*Section* 4.5): To prove Theorem 17, we need to bound all spectral radii of $B_{|K_0}$ for all weights a_i with uniformly bounded norm. We will use a net argument on the norm of $(B_{|K_0})^\ell$ conditioned on the event that there are no tangles in G^σ.

Remark 1. Let us comment on the main differences with [31], [10], [9], notably [9], which is the closest. The steps (i)–(iii) are similar to [9]. In the analog of step (iv), the work in [9] is greatly simplified by the fact that, with the terminology of the present paper, the weights a_i are matrices of the standard basis E_{uv}, $u, v \in [r]$. In this special case, the spectral radius of $\rho(B_\star)$ has an explicit combinatorial expression and products of a_i have a simple combinatorial description. Finally, step (v) is not present in [9].

4.2. *Path decomposition.* In this subsection, we set $X = [n]$ and let A be as in (1). We denote by B the non-backtracking matrix of A. Here we give an upper bound on $\rho(B_{|K_0})$ in terms of operator norms of new matrices that will be tuned for the use of the method of Füredi and Komlós. We fix a positive integer ℓ. The right-hand side of equation (27) can be studied by a weighted expansion of paths. To this end, we will use some definitions for the sequences in E that we will encounter to express the entries of B^ℓ as a weighted sum of non-backtracking paths. Recall the definition of a colored edge $[x, i, y]$ and a colored graph in Definition 6, and see Figure 4 for an illustration of the new definitions. Recall that $E = X \times [d]$.

Definition 18. For a positive integer k, let $\gamma = (\gamma_1, \ldots, \gamma_k) \in E^k$, with $\gamma_t = (x_t, i_t)$.

– The sets of vertices and pairs of colored edges of γ are the sets $V_\gamma = \{x_t : 1 \le t \le k\}$ and $E_\gamma = \{[x_t, i_t, x_{t+1}] : 1 \le t \le k-1\}$. We denote by G_γ the colored graph with vertex set V_γ and colored edges E_γ.
– An (*extended*) *path* of length $k-1$ is an element of E^k. The path γ is *non-backtracking* if for any $t \ge 1$, $i_{t+1} \ne i_t^*$. The subset of non-backtracking paths of E^k is Γ^k. If $e, f \in E$, we denote by Γ^k_{ef} paths in Γ^k such that $\gamma_1 = e$, $\gamma_k = f$.
– The *weights* of the path γ is the element of $M_r(\mathbb{C})$,

$$a(\gamma) = \prod_{t=2}^{k} a_{i_t}.$$

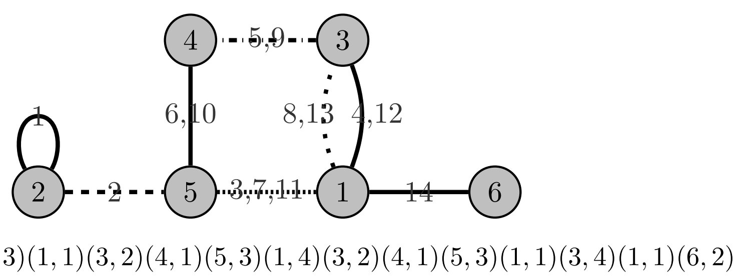


Figure 4. *An example where the involution i^* is the identity*: The colored graph G_γ is associated to a path $\gamma \in \Gamma^{15}$. The numbers on the edges are the values of t such that $[x_t, i_t, x_{t+1}]$ is equal to this edge. We have $V_\gamma = [6]$ and $E_\gamma = \{[2,1,2], [2,2,5], [5,3,1], [1,1,3], [3,2,4], [4,1,5], [1,4,3], [1,1,6]\}$.

By construction, from (12) we find that

$$(B^\ell)_{ef} = \sum_{\gamma \in \Gamma^{\ell+1}_{ef}} a(\gamma) \prod_{t=1}^{\ell} (S_{i_t})_{x_t x_{t+1}}.$$

Observe that in the above expression, $\Gamma^{\ell+1}$ and $a(\gamma)$ do not depend on the permutation matrices $S_i, i \in [d]$. We set

$$\underline{S}_i = S_i - \frac{1}{n}\mathbf{1} \otimes \mathbf{1}. \tag{28}$$

Note that $\underline{S}_i$ is the orthogonal projection of S_i on $\mathbf{1}^{\perp}$. Hence, setting $\underline{B} = \sum_{j \neq i^*} a_j \otimes \underline{S}_i \otimes E_{ij}$ as in (12), we get that, if $g \in K_0$, then

$$B^{\ell} g = \underline{B}^{\ell} g. \tag{29}$$

Moreover, arguing as above we find

$$(\underline{B}^{\ell})_{ef} = \sum_{\gamma \in \Gamma_{ef}^{\ell+1}} a(\gamma) \prod_{t=1}^{\ell} (\underline{S}_{i_t})_{x_t x_{t+1}}. \tag{30}$$

The matrix $\underline{B}$ will however not be used. Indeed, as pointed in Section 4.1, due to polynomially small events that would have had a big influence on the expected value of B^{ℓ} or $\underline{B}^{\ell}$ for large ℓ, we will first reduce the above sum over $\Gamma_{ef}^{\ell+1}$ to a sum over a smaller subset. We will only afterward project on K_0, which will create some extra remainder terms. We now introduce a key definition. (Recall the definition of cycles and $(H, x)_{\ell}$ in Definition 6.)

Definition 19 (Tangles). A graph H is *tangle-free* if it contains at most one cycle, and H is *ℓ-tangle-free* if for any vertex x, $(H, x)_{\ell}$ contains at most one cycle. Otherwise, H is tangled or ℓ-tangled. We say that $\gamma \in E^k$ is tangle-free or tangled if G_{γ} is. Finally, F^k and F_{ef}^k will denote the subsets of tangle-free paths in Γ^k and Γ_{ef}^k.

Now, recall the definition of the colored graph G^{σ} in Definition 6. Obviously, if G^{σ} is ℓ-tangle-free and $0 \leq k \leq 2\ell$, then

$$B^k = B^{(k)}, \tag{31}$$

where

$$(B^{(k)})_{ef} = \sum_{\gamma \in F_{ef}^{k+1}} a(\gamma) \prod_{t=1}^{k} (S_{i_t})_{x_t x_{t+1}}.$$

Similarly we define the matrix $\underline{B}^{(k)}$ by

$$(\underline{B}^{(k)})_{ef} = \sum_{\gamma \in F_{ef}^{k+1}} a(\gamma) \prod_{t=1}^{k} (\underline{S}_{i_t})_{x_t x_{t+1}}. \tag{32}$$

Beware that even if G^{σ} is ℓ-tangle-free and $2 \leq k \leq \ell$, $\underline{B}^k$ is a priori different from $\underline{B}^{(k)}$. (In (30) and (32) the summand is the same but the sum in (30) is over a larger set.) Nevertheless, at the cost of extra terms, as in (29), we may still express $B^{(\ell)}g$ in terms of $\underline{B}^{(\ell)}g$ for all $g \in K_0$. We start with the following telescopic sum decomposition:

$$(B^{(\ell)})_{ef} = (\underline{B}^{(\ell)})_{ef} + \sum_{\gamma \in F^{\ell+1}_{ef}} a(\gamma) \sum_{k=1}^{\ell} \left(\prod_{t=1}^{k-1} (\underline{S}_{i_t})_{x_t x_{t+1}} \right) \left(\frac{1}{n} \right) \left(\prod_{t=k+1}^{\ell} (S_{i_t})_{x_t x_{t+1}} \right), \tag{33}$$

which follows from the identity,

$$\prod_{t=1}^{\ell} x_t = \prod_{t=1}^{\ell} y_t + \sum_{k=1}^{\ell} \left(\prod_{t=1}^{k-1} y_t \right) (x_k - y_k) \left(\prod_{t=k+1}^{\ell} x_t \right).$$

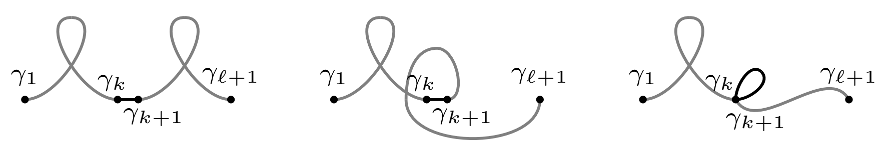


Figure 5. Tangle-free paths whose union is tangled.

We now rewrite (33) as a sum of matrix products for lower powers of $\underline{B}^{(k)}$ and $B^{(k)}$ up to some remainder terms. We decompose a path $\gamma = (\gamma_1, \ldots, \gamma_{\ell+1}) \in \Gamma^{\ell+1}$ as a path $\gamma' = (\gamma_1, \ldots, \gamma_k) \in \Gamma^k$, a path $\gamma'' = (\gamma_k, \gamma_{k+1}) \in \Gamma^2$ and a path $\gamma''' = (\gamma_{k+1}, \ldots, \gamma_{\ell+1}) \in \Gamma^{\ell-k+1}$. If the path γ is in $F^{\ell+1}$ (that is, γ tangle-free), then the three paths are tangle-free, but the converse is not necessarily true; see Figure 5. This will be the origin of the remainder terms. For $1 \le k \le \ell$, we denote by $F^{\ell+1}_k$ the set of $\gamma \in \Gamma^{\ell+1}$ as above such that $\gamma' \in F^k$, $\gamma'' \in F^2 = \Gamma^2$ and $\gamma''' \in F^{\ell-k+1}$. Then $F^{\ell+1} \subset F^{\ell+1}_k$. Setting, $F^{\ell+1}_{k,ef} = F^{\ell+1}_k \cap \Gamma^{\ell+1}_{ef}$, we write in (33)

$$\sum_{\gamma \in F^{\ell+1}_{ef}} (\star) = \sum_{\gamma \in F^{\ell+1}_{k,ef}} (\star) - \sum_{\gamma \in F^{\ell+1}_{k,ef} \backslash F^{\ell+1}_{ef}} (\star),$$

where $(\star)$ is the summand on the right-hand side of (33). We have

$$a(\gamma) = a(\gamma')a(\gamma'')a(\gamma''').$$

We denote by $\overline{B}$ the matrix on $\mathbb{C}^r \otimes \mathbb{C}^E$ defined by

$$\overline{B} = \sum_{j \neq i^*} a_j \otimes (\mathbf{1} \otimes \mathbf{1}) \otimes E_{ij}.$$

Observe that $\overline{B}_{ef} = \sum a(\gamma)$, where the sum is over all γ in $\Gamma^2_{ef} = F^2_{ef}$. We get

$$\sum_{\gamma \in F^{\ell+1}_{k,ef}} a(\gamma) \left(\prod_{t=1}^{k-1} (\underline{S}_{i_t})_{x_t x_{t+1}} \right) \left(\frac{1}{n} \right) \left(\prod_{t=k+1}^{\ell} (S_{i_t})_{x_t x_{t+1}} \right) = \left(\frac{1}{n} \right) \underline{B}^{(k-1)} \left(\overline{B} B^{(\ell-k)} \right)_{ef}.$$

For all $e, f \in E$, we set

$$(R^{(\ell)}_k)_{ef} = \sum_{\gamma \in F^{\ell+1}_{k,ef} \backslash F^{\ell+1}_{ef}} a(\gamma) \left(\prod_{t=1}^{k-1} (\underline{S}_{i_t})_{x_t x_{t+1}} \right) \left(\prod_{t=k+1}^{\ell} (S_{i_t})_{x_t x_{t+1}} \right). \tag{34}$$

We have from (33) that

$$B^{(\ell)} = \underline{B}^{(\ell)} + \frac{1}{n} \sum_{k=1}^{\ell} \underline{B}^{(k-1)} \overline{B} B^{(\ell-k)} - \frac{1}{n} \sum_{k=1}^{\ell} R^{(\ell)}_k.$$

Now, observe that if G^σ is ℓ-tangle free, then from (31), $\overline{B} B^{(\ell-k)} = \overline{B} B^{\ell-k}$. Moreover, the kernel of $\overline{B}$ contains K_0. Since $B^{\ell-k} K_0 \subset K_0$, we find that $\overline{B} B^{\ell-k} = 0$ on K_0. So finally, if G^σ is ℓ-tangle free, then for any $g \in K_0$,

$$B^\ell g = \underline{B}^{(\ell)} g - \frac{1}{n} \sum_{k=1}^{\ell} R^{(\ell)}_k g.$$

Putting this last inequality in (27), the outcome of this subsection is the following lemma.

LEMMA 20. *Let $\ell \geq 1$ be an integer, and let A be as in* (1) *be such that G^σ is ℓ-tangle free. Then,*

$$\rho(B_{|K_0}) \leq \left(\|\underline{B}^{(\ell)}\| + \frac{1}{n} \sum_{k=1}^{\ell} \|R^{(\ell)}_k\| \right)^{1/\ell}.$$

4.3. *Estimates on random permutations.* In this subsection we study some properties of permutations matrices S_i for the symmetric random permutations.

The first proposition gives a sharp estimate on the expected product of the variables $(\underline{S}_i)_{xy}$. This estimate will be used to bound entries in products of $\underline{B}^{(\ell)}, R^{(\ell)}_k$ and their transposes. Note that if $i \neq i^*$, then $(\underline{S}_i)_{xy}$ is centered: $\mathbb{E}(\underline{S}_i)_{xy} = 0$, while if $i = i^*$, then $(\underline{S}_i)_{xy}$ is almost centered, i.e., for $x \neq y$, $\mathbb{E}(\underline{S}_i)_{xy} = 1/(n-1) - 1/n = O(1/n^2)$.

We start with some definitions on colored graphs (as defined in Definition 6).

Definition 21.

- Let H be a colored graph with colored edge set E_H. A colored edge $e = [x, i, y] \in E_H$ is *consistent* if for any $e' = [x', i', y'] \in E_H$, $(x, i) = (x', i')$ or $(y, i^*) = (x', i')$ implies that $e = e'$. (Recall that $[x, i, y] = [y, i^*, x]$.) It is *inconsistent* otherwise.
- For a sequence of colored edges $(e_1, \dots, e_\tau)$, the *multiplicity* of $e \in \{e_t : 1 \leq t \leq \tau\}$ is $\sum_t \mathbf{1}(e_t = e)$. The edge e is consistent or inconsistent if it is consistent or inconsistent in the colored graph spanned by $\{e_t : 1 \leq t \leq \tau\}$.

In Figure 4, the edges $[1, 1, 6]$ and $[1, 1, 3]$ are inconsistent.

PROPOSITION 22 ([9]). *For symmetric random permutations, there exists a constant $c > 0$ such that for any sequence of colored edges $(f_1, \dots, f_\tau)$, with $f_t = [x_t, i_t, y_t]$, $\tau \leq \sqrt{n}$ and any $\tau_0 \leq \tau$, we have*

$$\left| \mathbb{E} \prod_{t=1}^{\tau_0} (\underline{S}_{i_t})_{x_t y_t} \prod_{t=\tau_0+1}^{\tau} (S_{i_t})_{x_t y_t} \right| \leq c\, 2^b \left(\frac{1}{n}\right)^e \left(\frac{3\tau}{\sqrt{n}}\right)^{e_1},$$

where $e = |\{f_t : 1 \leq t \leq \tau\}|$, b is the number of inconsistent edges and e_1 is the number of $1 \leq t \leq \tau_0$ such that $[x_t, i_t, x_{t+1}]$ is consistent and has multiplicity one.

Proof. Using the independence of the matrices S_i (up to the involution), the claim is contained in [9, Prop. 8] for matchings and [9, Prop. 25] for permutations. □

Recall that the graph G^σ is the colored graph with vertex set of $[n]$ and edges set of $[x, i, y]$ such that $\sigma_i(x) = y$ (and $\sigma_{i^*}(y) = x$).

LEMMA 23. *Let A be as in (1) for symmetric random permutation. For some constant $c > 0$, for any integer $1 \leq \ell \leq \sqrt{n}$, the expected number of cycles of length ℓ in G^σ is bounded by $c(d-1)^\ell$. The probability that G^σ is ℓ-tangled is at most $c\ell^3 (d-1)^{4\ell}/n$.*

Proof. Let H be a colored graph as in Definition 6, with vertex set $V_H \subset [n]$ and edge set E_H. Let us say that H is consistently colored if all its edges are consistent (as per Definition 21). If H is not consistently colored, then the probability that $H \subset G^\sigma$ is 0. Assume that H is consistently colored and that E_H contains e_i edges of the form $[x, i, y]$. If $i \neq i^*$, the probability these e_i edges are present in G^σ is

$$\prod_{t=0}^{e_i - 1} \frac{1}{n-t} \leq \left(\frac{1}{n - e_i + 1}\right)^{e_i}.$$

If $i = i^*$, this probability is

$$\prod_{t=0}^{e_i-1} \frac{1}{n-2t-1} \leq \left(\frac{1}{n-2e_i+1}\right)^{e_i}.$$

We use that, for any integers, k, ℓ with $k\ell \leq \alpha n$, $\ell \leq n/2$,

$$(n-\ell)^k \geq e^{-2\alpha} n^k.$$

(Indeed, $(n-\ell)^k = n^k \exp(k \log(1-\ell/n)) \geq n^k \exp(-2k\ell/n)$ since $\log(1-x) \geq -x/(1-x)$ for $0 \leq x < 1$.) Using the independence of the permutations σ_i (up to the involution), we deduce that, if $|E_H| \leq \sqrt{n}$, then

$$\mathbb{P}(H \subset G^\sigma) \leq c\left(\frac{1}{n}\right)^{|E_H|}, \tag{35}$$

for some constant $c > 0$.

Now, the number of properly consistently cycles in $[n]$ of length ℓ is at most

$$n^\ell d(d-1)^{\ell-1};$$

indeed, n^ℓ bounds the possible choices of the vertex set and $d(d-1)^{\ell-1}$ the possible colors of the edges. Since a cycle has ℓ edges, we get from (35) that the expected number of cycles of length ℓ contained in G^σ is at most $cd(d-1)^{\ell-1}$ as claimed.

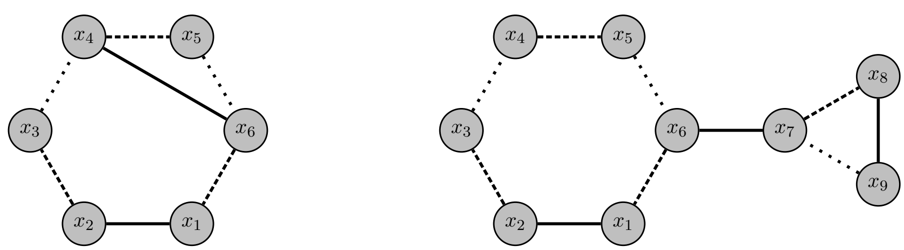


Figure 6. The involution i^* is the identity. On the left-hand side, we have a consistently colored $H_{6,1,4}$, and on the right-hand side, we have a consistently colored $H_{6,3,1}$.

Similarly, if G^σ is ℓ-tangled, then there exists a ball of radius ℓ that contains two cycles. Depending on whether these two cycles intersect or not, it follows that G^σ contains as a subgraph, either two cycles connected by a line segment or a cycle and a line segment (see Figure 6), where the line segment can be of length 0. More formally, for integers $1 \leq s \leq k$ and $m \geq 1$, define $H_{k,m,s}$ as the colored graph with vertex set $\{x_t : 1 \leq t \leq k+m-1\}$ of size $k+m-1$ and colored edges, for $1 \leq t \leq k+m-1$, $[x_t, i_t, x_{t+1}]$, with $x_{k+m} = x_s$ and $[x_k, i_{k+m}, x_1]$, where all $k+m$ edges are distinct. The graph

$H_{k,m,s}$ depends implicitly on the choice of the x_t's and i_t's. Similarly, for integers $k, k' \geq 1$ and $m \geq 0$, let $H'_{k,k',m}$ be the colored graph with vertex set $\{x_t : 1 \leq t \leq k+k'+m-1\}$ of size $k+k'+m-1$ and colored edges $[x_t, i_t, x_{t+1}]$ for $1 \leq t \leq k+k'+m-1$ with $x_{k+k'+m} = x_{k+m}$, and the edge $[x_k, i_{k+k'+m}, x_1]$. Again, the graph $H'_{k,k',m}$ depends implicitly on the choice of the x_t's and i_t's. Then, if G^σ is ℓ-tangled either it contains as a subgraph, for some x_t's and i_t's, a consistently colored graph $H_{k,m,s}$ with $k, m \leq 2\ell$ or a consistently colored graph $H_{k,k',m}$ with $k, k' + m \leq 2\ell$.

The number of consistently colored graphs $H_{k,m,s}$ in $[n]$ is at most

$$n^{k+m-1} d(d-1)^{k+m-1},$$

and the number of consistently colored graphs $H'_{k,k',m}$ in $[n]$ is at most

$$n^{k+k'+m-1} d(d-1)^{k+k'+m-1}.$$

From (35), we deduce that the probability that G^σ is ℓ-tangled is at most

$$\sum_{k,s,m \leq 2\ell} n^{k+m-1} d(d-1)^{k+m-1} c\left(\frac{1}{n}\right)^{k+m} + \sum_{k,k'+m \leq 2\ell} n^{k+k'+m-1} d(d-1)^{k+k'+m-1} c\left(\frac{1}{n}\right)^{k+k'+m}.$$

The latter is $O\left(\frac{\ell^3 (d-1)^{4\ell}}{n}\right)$ as claimed. □

4.4. *Trace method of Füredi and Komlós.*

4.4.1. *Norm of* $\underline{B}^{(\ell)}$. Here, we give a sharp bound on the operator norm of the matrices $\underline{B}^{(\ell)}$ for symmetric random permutations. In this subsection, we fix a collection $(a_i), i \in [d]$, of matrices such that the symmetry condition (3) holds and we assume $\max_i(\|a_i\| \vee \|a_i^{-1}\|^{-1}) \leq \varepsilon^{-1}$ for some $\varepsilon > 0$. Then $B_\star$ is the corresponding non-backtracking operator in the free group. The constants may depend implicitly on r, d and ε.

PROPOSITION 24. *Let* $\varepsilon > 0$. *If* $1 \leq \ell \leq \log n$, *then the event*

$$\|\underline{B}^{(\ell)}\| \leq (\log n)^{20} (\rho(B_\star) + \varepsilon)^\ell$$

holds with the probability at least $1 - ce^{-\frac{\ell \log n}{c \log \log n}}$, *where* $c > 0$ *depends on* r, d *and* ε.

The proof relies on the method of moments. Let m be a positive integer. With the convention that $f_{2m+1} = f_1$, we get

(36)
$$\begin{aligned}\|\underline{B}^{(\ell)}\|^{2m} = \|\underline{B}^{(\ell)}\underline{B}^{(\ell)*}\|^{m} &\leq \operatorname{tr}\Big\{\big(\underline{B}^{(\ell)}\underline{B}^{(\ell)*}\big)^{m}\Big\} \\ &= \sum_{(f_1,\ldots,f_{2m})\in E^{2m}} \operatorname{tr}\prod_{j=1}^{m}(\underline{B}^{(\ell)})_{f_{2j-1},f_{2j}}(\underline{B}^{(\ell)*})_{f_{2j},f_{2j+1}} \\ &= \sum_{\gamma\in W_{\ell,m}} \prod_{j=1}^{2m}\prod_{t=1}^{\ell}(\underline{S}_{i_{j,t}})_{x_{j,t}x_{j,t+1}} \operatorname{tr}\prod_{j=1}^{2m} a(\gamma_j)^{\varepsilon_j},\end{aligned}$$

where $a(\gamma_j)^{\varepsilon_j} = a(\gamma_j)$ or $a(\gamma_j)^*$ depending on the parity of j and $W_{\ell,m}$ is the set of $\gamma = (\gamma_1,\ldots,\gamma_{2m})$ such that $\gamma_j = (\gamma_{j,1},\ldots,\gamma_{j,\ell+1}) \in F^{\ell+1}$, $\gamma_{j,t} = (x_{j,t}, i_{j,t})$, and for all $j = 1,\ldots,m$,

$$\gamma_{2j,1} = \gamma_{2j+1,1} \quad \text{and} \quad \gamma_{2j-1,\ell+1} = \gamma_{2j,\ell+1}, \tag{37}$$

with the convention that $\gamma_{2m+1} = \gamma_1$; see Figure 7. The proof of Proposition 24 is based on an upper bound on the expectation of the right-hand side of (36). We write

$$\mathbb{E}\|\underline{B}^{(\ell)}\|^{2m} \leq \sum_{\gamma\in W_{\ell,m}} |w(\gamma)| \operatorname{tr}|a(\gamma)|, \tag{38}$$

where we have set

$$w(\gamma) = \mathbb{E}\prod_{t=1}^{\ell}(\underline{S}_{i_{j,t}})_{x_{j,t}x_{j,t+1}} \quad \text{and} \quad a(\gamma) = \prod_{j=1}^{2m} a(\gamma_j)^{\varepsilon_j}.$$

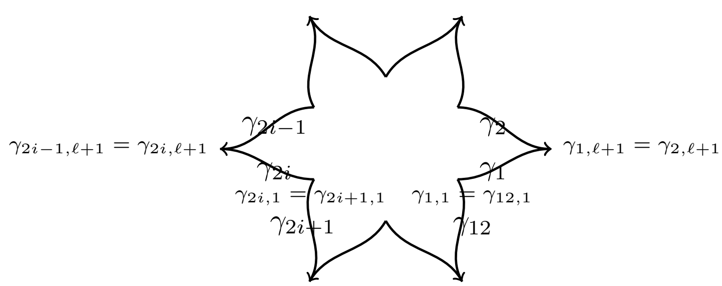


Figure 7. A path $\gamma = (\gamma_1,\ldots,\gamma_{12})$ in $W_{\ell,6}$; each γ_i is tangle-free.

First, to deal with this large sum, we partition $W_{\ell,m}$ in isomorphism classes. Permutations on $[n]$ and $[d]$ act naturally on $W_{\ell,m}$. We consider the isomorphism class $\gamma \sim \gamma'$ if there exist $\sigma \in \mathcal{S}_n$ and $(\tau_x)_x \in (\mathcal{S}_d)^n$ such that, with $\gamma'_{j,t} = (x'_{j,t}, i'_{j,t})$, for all $1 \leq j \leq 2m$, $1 \leq t \leq \ell+1$, $x'_{j,t} = \sigma(x_{j,t})$,

$i'_{j,t} = \tau_{x_{j,t}}(i_{j,t})$ and $(i'_{j,t})^* = \tau_{x_{j,t+1}}((i_{j,t})^*)$. For each $\gamma \in W_{\ell,m}$, we define G_γ as in Definition 18: $V_\gamma = \cup_j V_{\gamma_j} = \{x_{j,t} : 1 \le j \le 2m, 1 \le t \le \ell+1\}$ and $E_\gamma = \cup_j E_{\gamma_j} = \{[x_{j,t}, i_{j,t}, x_{j,t+1}] : 1 \le j \le 2m, 1 \le t \le \ell\}$ are the sets of visited vertices and visited pairs of colored edges along the path. Importantly, G_γ is connected. We may then define a canonical element in each isomorphic class as follows. We say that a path $\gamma \in W_{\ell,m}$ is *canonical* if γ is minimal in its isomorphism class for the lexicographic order (x before $x+1$ and (x,i) before $(x,i+1)$), that is, $\gamma_{1,1} = (1,1)$ and $\gamma_{j,t}$ minimal over all $\gamma'_{j,t}$ such that $\gamma' \sim \gamma$ and $\gamma'_{k,s} = \gamma_{k,s}$ for all $(k,s) \prec (j,t)$. Our first lemma bounds the number of isomorphism classes. This lemma is a variant of [10, Lemma 17] and [9, Lemma 13]. It relies crucially on the fact that an element $\gamma \in W_{\ell,m}$ is composed of $2m$ tangle-free paths.

LEMMA 25. *Let $\mathcal{W}_{\ell,m}(v,e)$ be the subset of canonical paths with $|V_\gamma| = v$ and $|E_\gamma| = e$. We have*

$$|\mathcal{W}_{\ell,m}(v,e)| \le (2d\ell m)^{6m\chi+10m},$$

with $\chi = e - v + 1 \ge 0$.

Proof. We bound $|\mathcal{W}_{\ell,m}(v,e)|$ by constructing an encoding of the canonical paths (that is, an injective map from $\mathcal{W}_{\ell,m}(v,e)$ to a set whose cardinality is easily upper bounded). For $i \le i \le 2m$ and $1 \le t \le \ell$, let $e_{j,t} = (x_{j,t}, i_{j,t}, x_{j,t+1})$ and $[e_{j,t}] = [x_{j,t}, i_{j,t}, x_{j,t+1}] \in E_\gamma$ be the corresponding colored edge. We explore the sequence $(e_{i,t})$ in lexicographic order denoted by $\preceq$ (that is, $(j,t) \preceq (j+1,t')$ and $(j,t) \preceq (j,t+1)$). We think of the index (j,t) as a time. We define $(j,t)^-$ as the largest time smaller than (j,t), i.e., $(j,t)^- = (j,t-1)$ if $t \ge 2$, $(j,1)^- = (j-1,\ell)$ if $j \ge 2$ and, by convention, $(1,1)^- = (1,0)$.

We denote by $G_{(j,t)}$ the graph spanned by the edges $\{[e_{j',t'}] : (j',t') \preceq (j,t)\}$. The graphs $G_{(j,t)}$ are non-decreasing over time and by definition $G_{(2m,\ell)} = G_\gamma$. We may define a growing spanning forest $T_{(j,t)}$ of $G_{(j,t)}$ as follows: $T_{(1,0)}$ has no edge and a single vertex, 1. Then, $T_{(j,t)}$ is obtained from $T_{(j,t)^-}$ by adding the edge $[e_{j,t}]$ if its addition does not create a cycle in $T_{(j,t)^-}$. We then say that $[e_{j,t}]$ is a *tree edge*. By construction $T_{(j,t)}$ is a spanning forest of $G_{(j,t)}$ and $T_\gamma = T_{(2m,\ell)}$ is a spanning tree of G_γ. An edge $[e_{j,t}]$ in $G_\gamma \backslash T_\gamma$ is called an *excess edge*. Since T_γ has $v-1$ edges, we have

$$\chi = |\{f \in E_\gamma : \ f \text{ is an excess edge}\}| = e - v + 1 \ge 0. \tag{39}$$

Now, from (37), for each j, there is a smallest time (j,σ), which we call *the merging time*, such that $G_{j,\sigma}$ will be connected. By convention, if $x_{j,1} \in G_{(j,1)^-}$, we set $\sigma = 0$. (For example from (37), if j is odd, then $\sigma = 0$.) We say that (j,t) is a *first time* if it is not a merging time and if $[e_{j,t}]$ is a tree edge that has not been seen before (that is, $e_{j,t} \neq e_{k,s}$ for all $(k,s) \preceq (j,t)$). We say that (j,t) is an *important time* if $[e_{j,t}]$ is an excess edge (see Figure 8).

$$\gamma_1 = (1,1)(2,2)(3,1)(4,3)(2,2)(3,1)(4,3)(2,2)(3,1)(4,3)(2,2)(3,1)(5,2)$$

Figure 8. *A canonical path* $\gamma_1 \in F^{13}$ *and its associated spanning tree*; *the involution* i^* *is the identity*: The times $(1,t)$ with $t \in \{1,2,3,12\}$ are first times and $t = \{4,7,10\}$ are important times, $(1,4)$ is the short cycling time, and $(1,7),(1,10)$ are superfluous. With the notation below, $t_1 = 4$, $t_0 = 2$, $t_2 = 12$, and $\tau = 13$.

By construction, since the path γ_j is non-backtracking, it can be decomposed by the successive repetition of (i) a sequence of first times (possibly empty), (ii) an important time or a merging time, and (iii) a path on the forest defined so far (possibly empty). Note also that if $t \geq 2$ and (j,t) is a first time, then $i_t = p$ and $x_{j,t+1} = m+1$, where m is the number of previous first times (including (j,t)) and p is minimal over all $i \geq 1$, such that $i \neq i^*_{t-1}$. Indeed, since γ is canonical, every time that a new vertex in V_γ is visited its number has to be minimal, and similarly for the number of the color of a new edge. It follows that if $(j,t),\dots,(j,t+s)$ are first times and $x_{j,t}$ is known, then the values of $e_{j,t},\dots,e_{j,t+s}$ can be unambiguously computed.

We can now build a first encoding of $\mathcal{W}_{\ell,m}$. If (j,t) is an important time, we mark the time (j,t) by the vector $(i_{j,t}, x_{j,t+1}, x_{j,\tau}, i_{j,\tau})$, where (j,τ) is the next time that $e_{j,\tau}$ will not be a tree edge of the forest $T_{j,t}$ constructed so far. (By convention, if the path γ_j remains on the tree, we set $\tau = \ell+1$.) For $t = 1$, we also add the *starting mark* $(x_{j,1}, \sigma, x_{j,\tau}, i_{j,\tau})$, where σ is the merging time and $(j,\tau) \geq (j,\sigma)$ is as above the next time that $[e_{j,\tau}]$ will not be a tree edge of the forest constructed so far. Since there is a unique non-backtracking path between two vertices of a tree, we can reconstruct $\gamma \in \mathcal{W}_{\ell,m}$ from the starting marks and the position of the important times and their marks. It gives rise to a first encoding.

In this encoding, the number of important times could be large (see Figure 8). We will now use the assumption that each path γ_j is tangle-free to partition important times into three categories, *short cycling*, *long cycling* and *superfluous* times. For each $j \in [2m]$, we consider the first occurrence of a time (j,t_1) such that $x_{j,t_1+1} \in \{x_{j,1},\dots,x_{j,t_1}\}$. If such t_1 exists, the last important time $(j,t_s) \preceq (j,t_1)$ will be called the short cycling time. Let $1 \leq t_0 \leq t_1$ be such that $x_{j,t_0} = x_{j,t_1+1}$. By assumption, $C_j = (e_{j,t_0},\dots,e_{j,t_1-1})$ will be the unique cycle visited by γ_j. We denote by (j, t_2) the next $t_2 \geq t_1$ such that e_{j,t_2} in not in C_j. (By convention, $t_2 = \ell + 1$ if γ_j remains in C_j.) We modify the mark of the short cycling time (j, t_s) as $(i_{j,t_s}, x_{j,t_s+1}, x_{j,t_1}, t_2, x_{j,\tau}, i_{j,\tau})$, where $(j, \tau) \succeq (j, t_2)$ is the next time that $[e_{j,\tau}]$ will not be a tree edge of the forest constructed so far. Important times (j, t) with $1 \leq t < t_s$ or $\tau \leq t \leq \ell$ are called long cycling times; they receive the usual mark $(i_{j,t}, x_{j,t+1}, x_{j,\tau}, i_{j,\tau})$. The other important times are called superfluous. By convention, if there is no short cycling time, we call anyway, the last important time, the short cycling time. We observe that for each j, the number of long cycling times on γ_j is bounded by $\chi - 1$. (Since there is at most one cycle, no edge of E_γ can be seen twice outside those of C_j, the -1 coming from the fact the short cycling time is an excess edge.)

We now have our second encoding. We can reconstruct γ from the starting marks, the positions of the long cycling and the short cycling times and their marks. For each j, there are at most 1 short cycling time and $\chi - 1$ long cycling times. There are at most $(\ell+1)^{2m\chi}$ ways to position them. There are at most $d^2 v^2$ different possible marks for a long cycling time and $d^2 v^3(\ell + 1)$ possible marks for a short cycling time. Finally, there are $dv^2(\ell + 1)$ possibilities for a starting mark. We deduce that

$$|\mathcal{W}_{\ell,m}(v, e)| \leq (\ell+1)^{4m\chi}(dv^2(\ell + 1))^{4m}(d^2v^2)^{4m(\chi-1)}(d^2v^3(\ell + 1))^{4m}.$$

Using $v \leq 2\ell m + 1$ and $\ell + 1 \leq 2\ell$, we obtain the claimed bound. □

Our second lemma bounds the sum of $a(\gamma)$ in an equivalence class.

LEMMA 26. *Let $\rho = \rho(B_\star)+\varepsilon$ and k_0 be a positive integer. Then, there exists a constant $c > 0$ depending on r, d and ε such that for any $\gamma \in \mathcal{W}_{\ell,m}(v, e)$,*

$$\sum_{\gamma' : \gamma' \sim \gamma} \operatorname{tr}|a(\gamma')| \leq c^{m+\chi+e_1} n^v \rho_0^{2(\ell m - v)} \rho^{2v}, \tag{40}$$

where $\chi = e - v + 1$, e_1 is the number of edges of E_γ with multiplicity one and

$$\rho_0 = \max \left\| \prod_{s=1}^{k_0} a_{i_s} \right\|^{\frac{1}{k_0}},$$

and the maximum is over all non-backtracking sequences $(i_1, \ldots, i_{k_0})$, that is, $i_{s+1} \neq i_{s^*}$. Moreover, for all k_0 large enough, we have $\rho_0 \leq \rho$.

Proof. We start by proving (40). The proof relies on a decomposition of G_γ where the path is split into sub-paths on the free group. Let v_k (respectively $v_{\geq k}$) be the set of vertices of G_γ of degree k (respectively $\geq k$). We have

$$v_1 + v_2 + v_{\geq 3} = v \quad \text{and} \quad v_1 + 2v_2 + 3v_{\geq 3} \leq \sum_k k v_k = 2e.$$

Subtracting from the right-hand side twice the left-hand side, we deduce that

$$v_{\geq 3} \leq 2(e - v) + v_1 \leq 2\chi + 2m - 2.$$

Indeed, at the last step the bound $v_1 \leq 2m$ follows from the observation that since each γ_j is non-backtracking, only a vertex $x \in V_\gamma$ such that $x = x_{j,1}$ or $x = x_{j,\ell+1}$ for some $1 \leq j \leq 2m$ can be of degree 1.

Now, consider the set $V'_\gamma \subset V_\gamma$ formed by vertices of degree at least 3 and vertices $x \in V_\gamma$ such that $x = x_{i,1}$ or $x = x_{i,\ell+1}$ for some $1 \leq i \leq 2m$. From what precedes,

$$v' = |V'_\gamma| \leq 2\chi + 4m - 2.$$

We now build the graph G'_γ on V'_γ obtained from G_γ by merging degree 2 vertices along edges. More formally, let P_γ be the set of non-backtracking sequences $\pi = (y_1, i_1, \dots, y_k, i_k, y_{k+1})$ with $[y_s, i_s, y_{s+1}] \in E_\gamma$ for $1 \leq s \leq k$ and $y_1, y_{k+1} \in V'_\gamma$, $y_2, \dots, y_k \in V_\gamma \backslash V'_\gamma$. We set $\pi^* = (y_{k+1}, i^*_k, y_k, \dots, y_1) \in P_\gamma$. Since all vertices not in V'_γ have degree 2, two distinct paths $\pi, \pi' \in P_\gamma$ are either disjoint (except the endpoints) or $\pi^* = \pi'$. As in Definition 6, we define a (generalized) colored edge as an equivalence class $[\pi]$ of π in P_γ endowed with the equivalence $\pi \sim \pi'$ if $\pi' \in \{\pi, \pi^*\}$. Then $G'_\gamma = (V'_\gamma, E'_\gamma)$ is the colored graph with edge set E'_γ, the set of $[\pi]$ with $\pi = (y_1, i_1, \dots, y_k, i_k, y_{k+1}) \in P_\gamma$, $[\pi]$ being an edge between y_1 and y_{k+1}; see Figure 9. Let $e' = |E'_\gamma|$. We find easily that this operation of merging degree 2 vertices preserves the Euler characteristic. (If $\pi = (y_1, i_1, \dots, y_k, i_k, y_{k+1})$ is in P_γ, it replaces k edges and $k - 1$ vertices of G_γ by a single edge in G'_γ.) That is,

$$e' - v' + 1 = e - v + 1 = \chi.$$

It follows that

$$e' \leq 3\chi + 4m - 3. \tag{41}$$

Now, we recall the multiplicity introduced above Proposition 22. If $[x, i, y] \in E_\gamma$, the *multiplicity* of $[x, i, y]$, denoted by $m_{[x,i,y]}$, is the number of times that $[\gamma_{j,s}, i_{j,s}, \gamma_{j,s+1}] = [x, i, y]$. Since γ is non-backtracking along each edge of E'_γ, we observe that if $[\pi] \in E'_\gamma$ with $\pi = (y_1, i_1, \dots, y_k, i_k, y_{k+1})$, then all edges $[y_s, i_s, y_{s+1}]$ have the same multiplicity. We may thus unambiguously define the multiplicity $m_{[\pi]}$ of an edge $[\pi] \in E'_\gamma$; see Figure 9. Due to the symmetry condition (3), we note that the norm of product of a_i's along an edge $[\pi]$ does not depend on whether we take the product along π or π^*:

$$\left\| \prod_{s=1}^{k} a_{i_s} \right\| = \left\| \left(\prod_{s=1}^{k} a_{i_s} \right)^* \right\| = \left\| \prod_{s=1}^{k} a_{i^*_{k-s+1}} \right\|. \tag{42}$$

Let e_t be the number of edges of multiplicity equal to t. We have

$$\sum_t e_t = e \quad \text{and} \quad \sum_t t e_t = 2\ell m. \tag{43}$$

$$\gamma_1 = (1,1)(2,2)(3,1)(4,3)(2,2)(3,1)(5,2)$$
$$\gamma_2 = (1,1)(2,2)(2,3)(4,1)(3,2)(3,1)(5,2)$$

Figure 9. *A canonical path* $\gamma = (\gamma_1, \gamma_2) \in \mathcal{W}_{6,1}$ *and its associated graphs* G_γ *and* G'_γ; *the involution* i^* *is the identity*: We have that $V_\gamma \backslash V'_\gamma = \{4\}$ and that $[\pi]$ is an edge of G'_γ with $\pi = (2,3,4,1,3)$. This edge has multiplicity 2. The edge $[(1,1,2)]$ has multiplicity 2.

We find

$$\sum_t (t-2)_+ e_t = \sum_t (t-2)e_t + e_1 = 2(\ell m - e) + e_1. \tag{44}$$

Since the path γ_j is non-backtracking, we may decompose it into successive visits of the edges of E'_γ. More precisely, we decompose γ_j as $\gamma_j = (p_{j,1}, p_{j,2}, \ldots, p_{j,k_j})$ where either (i) $p_{j,t}$ follows an edge of E'_γ that is visited for the first or second time, or (ii) $p_{j,t}$ follows a sequence of edges of E'_γ that have been visited previously at least two times. By construction, in the decomposition of the whole path γ, there are at most $2e'$ subpaths $p_{j,t}$ of type (i) and thus $2e' + 4m$ subpaths of type (ii). We may then write

$$\begin{aligned} \|a(\gamma)\| &\leq \prod_{j=1}^{2m} \|a(\gamma_j)\| \\ &\leq \delta \prod_{p_{j,t}} \left\| \prod_{s=1}^{k} a_{i_s} \right\| = \delta \prod_{p_{j,t} \text{ type (i)}} \left\| \prod_{s=1}^{k} a_{i_s} \right\| \prod_{p_{j,t} \text{ type (ii)}} \left\| \prod_{s=1}^{k} a_{i_s} \right\|, \end{aligned} \tag{45}$$

where in the above product, $p_{j,t} = (y_1, i_1, y_2, i_2, \ldots, y_k, i_k)$ and

$$\delta = \prod_{j=1}^{2m} \frac{\|a_{i_{j,\ell+1}}\|}{\|a_{i_{j,1}}\|} \leq \varepsilon^{-4m}$$

accounts for the boundary effects. To estimate (45), we shall use the two rough bounds

$$\left\| \prod_{s=1}^{k} a_{i_s} \right\| \leq \left\| \prod_{s=1}^{k} a_{i_s} \right\|^2 \left\| \left(\prod_{s=1}^{k} a_{i_s} \right)^{-1} \right\| \leq \varepsilon^{-k} \left\| \prod_{s=1}^{k} a_{i_s} \right\|^2. \tag{46}$$

We notice also that since $\max_i \|a_i\| \leq \varepsilon^{-2}\rho_0$,

$$\left\|\prod_{s=1}^{k} a_{i_s}\right\| \leq \left(\prod_{s=1}^{\lfloor k/k_0 \rfloor} \|a_{i_{k_0 s - k_0+1}} \cdots a_{i_{k_0 s}}\|\right) \|a_{i_{k_0 \lfloor k/k_0 \rfloor +1}} \cdots a_{i_k}\| \tag{47}$$
$$\leq \rho_0^{\lfloor k/k_0 \rfloor} \max_i \|a_i\|^{k - k_0 \lfloor k/k_0 \rfloor} \leq \varepsilon^{-2k_0} \rho_0^k,$$

which uses the non-backtracking condition $i_{s+1} \neq i_s^*$. Now, in (45), we decompose the product over $p_{j,t}$ of type (i) and of type (ii). Using (42) and (46) when $m_{[\pi]} = 1$, we arrive at

$$\prod_{p_{j,t} \text{ type (i)}} \left\|\prod_{s=1}^{k} a_{i_s}\right\| = \prod_{[\pi] \in E'_\gamma} \left\|\prod_{s=1}^{k} a_{i_s}\right\|^{m_{[\pi]} \wedge 2}$$
$$\leq \prod_{[\pi] \in E'_\gamma} (\varepsilon^{-k})^{\mathbf{1}(m_{[\pi]}=1)} \left\|\prod_{s=1}^{k} a_{i_s}\right\|^2 = \varepsilon^{-e_1} \prod_{[\pi] \in E'_\gamma} \left\|\prod_{s=1}^{k} a_{i_s}\right\|^2,$$

where in the above product $\pi = (y_1, i_1, \ldots, y_k, i_k, y_{k+1})$. Similarly, for each $p_{j,t}$ of type (ii), we use (47). Since there are at most $2e' + 4m$ subpaths $p_{j,t}$ of type (ii) and since the sum of length of $p_{j,t}$ is equal to $\sum_t (t-2)_+ e_t = 2(\ell m - e) + e_1$ by (44), from (47), we find

$$\prod_{p_{j,t} \text{ type (ii)}} \left\|\prod_{s=1}^{k} a_{i_s}\right\| \leq \prod_{p_{j,t} \text{ type (ii)}} \varepsilon^{-2k_0} \rho_0^k \leq \varepsilon^{-2k_0(2e'+4m)} \rho_0^{2(\ell m - e) + e_1}.$$

We finally plug the last two upper bounds into (45). Using (41), for some $c > 0$, we arrive at

$$\|a(\gamma)\| \leq c^{m+\chi+e_1} \rho_0^{2(\ell m - e)} \prod_{[\pi] \in E'_\gamma} \left\|\prod_{s=1}^{k} a_{i_s}\right\|^2.$$

Thus, summing over all $\gamma' \sim \gamma$, we obtain

$$\sum_{\gamma' : \gamma' \sim \gamma} \|a(\gamma')\| \leq c^{m+\chi+e_1} \rho_0^{2(\ell m - e)} n^v \prod_{[\pi] \in E'_\gamma} \left(\sum \left\|\prod_{s=1}^{k} a_{i_s}\right\|^2\right), \tag{48}$$

where for $\pi = (y_1, i_1, \ldots, y_k, i_k, y_{k+1})$, the sum is over all non-backtracking sequence $(i_1, \ldots, i_k)$.

Finally, in this last expression, $\sum \left\|\prod_{s=1}^{k} a_{i_s}\right\|^2$ can be bounded in terms of the spectral radius of the non-backtracking on the free group. Let $B_\star$ be the non-backtracking operator on the free group associated to $A_\star$ defined in (4).

There exists $c > 0$ such that for any $(g, i) \in X \times [d]$, for any integer $k \geq 0$,

$$\|B_\star^k \delta_{(o,i)}\|_2^2 = \sum \left\| \prod_{s=2}^{k+1} a_{i_s} \right\|^2 \leq \max(c, \rho^{2k}),$$

where the sum is over all non-backtracking sequence $(i_1, \ldots, i_{k+1})$ such that $i_1 = i$. Moreover from Lemma 14, the same constant c may be taken for all $B_\star$ with weights such that $\max_i(\|a_i\|) \leq \varepsilon^{-1}$. Also, at the cost of changing the constant c and taking k_0 large enough, in (48) we find

$$\sum \left\| \prod_{s=1}^{k} a_{i_s} \right\|^2 \leq c\rho^{2k} \quad \text{and} \quad \sum \left\| \prod_{s=1}^{k_0} a_{i_s} \right\|^2 \leq \rho^{2k_0}. \tag{49}$$

Since the sum of the length of all $[\pi] \in E'_\gamma$ is e, from (48) we get

$$\sum_{\gamma' : \gamma' \sim \gamma} \|a(\gamma')\| \leq c^{m+\chi+e_1} \rho_0^{2(\ell m - e)} n^v c^{e'} \rho^{2e}.$$

It remains to again use (41), $e = v + \chi - 1$ and adjust the constant c. Since $\mathrm{tr}|a(\gamma)| \leq d\|a(\gamma)\|$, we obtain (40). Finally, the claimed lower bound, $\rho \geq \rho_0$ is a direct consequence of the right-hand side of (49) by considering only the non-backtracking sequence in the sum that maximizes $\left\|\prod_{s=1}^{k_0} a_{i_s}\right\|$. □

Our final lemma gives a bound on $w(\gamma)$ defined below (38). Observe that $w(\gamma)$ is invariant on each isomorphism class. In the sequel, for an integer $n \in \mathbb{Z}$, we set n_+ to be its positive part, i.e., $n_+ = \max(0, n)$.

LEMMA 27. *There exists a constant $c > 0$ such that for any $\gamma \in \mathcal{W}_{\ell,m}(v, e)$ and $2\ell m \leq \sqrt{n}$,*

$$|w(\gamma)| \leq c^{m+\chi} \left(\frac{1}{n}\right)^e \left(\frac{6\ell m}{\sqrt{n}}\right)^{(e_1 - 4\chi - 4m)_+},$$

with $\chi = e - v + 1$ and e_1 is the number of edges of E_γ with multiplicity one. Moreover,

$$e_1 \geq 2(e - \ell m).$$

Proof. We start by the last statement. Let $e_{\geq 2}$ be the number of edges of E_γ of multiplicity at least 2. From (43), we have

$$e_1 + e_{\geq 2} = e \qquad \text{and} \quad e_1 + 2e_{\geq 2} \leq 2\ell m.$$

Therefore, $e_1 \geq 2(e - \ell m)$ as claimed. Let b the number of inconsistent edges. (Recall the definition above Proposition 22.) Using the terminology of the proof of Lemma 25, a new inconsistent edge can appear at the the start of a sequence of first times, at a first visit of an excess edge or at the merging time. Every such step can create two inconsistent edges. Since each non-empty sequence of first times is followed either by a merging time or by a first visit of an excess edge, we deduce from (39) that $b \leq 4\chi + 4m$. So finally, the number of consistent edges of γ of multiplicity one is at least $(e_1 - 4\chi - 4m)_+$. It remains to apply Proposition 22. $\square$

All ingredients are in order to prove Proposition 24.

Proof of Proposition 24. For $n \geq 3$, we define

$$m = \left\lfloor \frac{\log n}{13 \log(\log n)} \right\rfloor . \tag{50}$$

For this choice of m, $n^{1/(2m)} = o(\log n)^7$ and $\ell m = o(\log n)^2$. Set $\rho = \rho(B_\star) + \varepsilon/2$. It suffices to prove that

$$S = \sum_{\gamma \in W_{\ell,m}} |w(\gamma)| \mathrm{tr}|a(\gamma)| \leq n(c\ell m)^{10m} \rho^{\ell m}. \tag{51}$$

Indeed, Proposition 24 follows immediately from (38)–(51) and Markov inequality. Recall that G_γ is connected for any $\gamma \in W_{\ell,m}$. Hence, $|E_\gamma| \geq |V_\gamma| - 1$ and

$$S \leq \sum_{v=1}^{\infty} \sum_{e=v-1}^{\infty} |\mathcal{W}_{\ell,m}(v,e)| \max_{\gamma \in \mathcal{W}_{\ell,m}(v,e)} \left(|w(\gamma)| \sum_{\gamma' \sim \gamma} \mathrm{tr}|a(\gamma')| \right).$$

Let $\gamma \in \mathcal{W}_{\ell,m}(v,e)$ with e_1 edges of multiplicity one and $\chi = e - v + 1$. We use Lemma 26 with $\varepsilon' = \varepsilon/2$ and Lemma 27. Since $a \leq b + (a-b)_+$ and $e_1 \geq 2(e - \ell m)$ (by Lemma 27), we find

$$\begin{aligned} |w(\gamma)| \sum_{\gamma' \sim \gamma} \mathrm{tr}|a(\gamma')| &\leq n^v c^{m+\chi+e_1} \rho_0^{2(\ell m - v)} \rho^{2v} \left(\frac{1}{n}\right)^v \left(\frac{6\ell m}{\sqrt{n}}\right)^{(e_1 - 4\chi - 4m)_+} \\ &\leq n c^{5(m+\chi)} \rho_0^{2(\ell m - v)} \rho^{2v} \left(\frac{1}{n}\right)^\chi \left(\frac{6c\ell m}{\sqrt{n}}\right)^{(2(v - \ell m - 1) - 2\chi - 4m)_+} . \end{aligned}$$

We set $\alpha = (8c\ell m)^2/n$ and $\ell' = \ell + 2$. Since $\rho \geq \rho_0$ (if k_0 is chosen large enough in Lemma 26), we deduce from Lemma 25 that, for some new constant $c > 0$,

$$\begin{aligned} S &\leq \sum_{v=1}^{\infty} \sum_{\chi=0}^{\infty} n(c\ell m)^{6m\chi+10m} \rho^{2\ell m} \left(\frac{1}{n}\right)^\chi \alpha^{(v - \ell' m - 1 - \chi)_+}. \\ &= S_1 + S_2 + S_3, \end{aligned}$$

where S_1 is the sum over $\{1 \le v \le \ell' m, \chi \ge 0\}$, S_2 over $\{v > \ell' m, 0 \le \chi < v - \ell' m\}$, and S_3 over $\{v > \ell' m, \chi \ge v - \ell' m\}$. We find,

$$\begin{aligned} S_1 &= n(c\ell m)^{10m} \rho^{2\ell m} \sum_{v=1}^{\ell' m} \sum_{\chi=0}^{\infty} \left(\frac{(c\ell m)^{6m}}{n} \right)^{\chi} \\ &\le n(c\ell m)^{10m} (\ell' m) \rho^{2\ell m} \sum_{\chi=0}^{\infty} \left(\frac{(c\ell m)^{6m}}{n} \right)^{\chi}. \end{aligned}$$

For our choice of m in (50), for n large enough,

$$\frac{(c\ell m)^{6m}}{n} \le \frac{(\log n)^{12m}}{n} \le n^{-1/13}.$$

In particular, the above geometric series converges and, adjusting the value of c, the right-hand side of (51) is an upper bound for S_1 (since $\ell' m \le c^m$ for $c > 1$ and n large enough). Similarly, since $\alpha = (8\ell m)^2/n$, for n large enough,

$$\begin{aligned} S_2 &= n(c\ell m)^{10m} \rho^{2\ell m} \sum_{v=\ell' m+1}^{\infty} \alpha^{v-\ell' m-1} \sum_{\chi=0}^{v-\ell' m-1} \left(\frac{(c\ell m)^{6m}}{\alpha n} \right)^{\chi} \\ &\le n(c\ell m)^{10m} \rho^{2\ell m} \sum_{v=\ell' m+1}^{\infty} \alpha^{v-\ell' m-1}\, 2 \left(\frac{(c\ell m)^{6m}}{\alpha n} \right)^{v-\ell' m-1} \\ &= 2n(c\ell m)^{10m} \rho^{2\ell m} \sum_{t=0}^{\infty} \left(\frac{(c\ell m)^{6m}}{n} \right)^{t}. \end{aligned}$$

Again, the geometric series is convergent and the right-hand side of (51) is an upper bound for S_2. Finally, the same manipulation gives for n large enough,

$$\begin{aligned} S_3 &= n(c\ell m)^{10m} \rho^{2\ell m} \sum_{v=\ell' m+1}^{\infty} \sum_{\chi=v-\ell' m}^{\infty} \left(\frac{(c\ell m)^{6m}}{n} \right)^{\chi} \\ &\le n(c\ell m)^{10m} \rho^{2\ell m} \sum_{v=\ell' m+1}^{\infty} 2 \left(\frac{(c\ell m)^{6m}}{n} \right)^{v-\ell' m} \\ &= 2n(c\ell m)^{10m} \rho^{2\ell m} \sum_{t=1}^{\infty} \left(\frac{(c\ell m)^{6m}}{n} \right)^{t}. \end{aligned}$$

The right-hand side of (51) is again an upper bound for S_3. It concludes the proof. □

4.4.2. *Norm of* $R_k^{(\ell)}$. Here, we give a rough bound on the operator norm of the matrices $R_k^{(\ell)}$ for symmetric random permutations. In this subsection, we fix a collection $(a_i), i \in [d]$, of matrices such that the symmetry condition (3) holds and we assume $\max_i(\|a_i\|) \le \varepsilon^{-1}$ for some $\varepsilon > 0$. The constants may depend implicitly on r, d and ε.

PROPOSITION 28. *For any* $1 \leq k, \ell \leq \log n$, *the event*

$$\|R_k^{(\ell)}\| \leq (\log n)^{40} \rho_1^\ell$$

holds with the probability at least $1 - ce^{-\frac{\ell \log n}{c \log\log n}}$, *where* $c > 0$ *and* $\rho_1 > 0$ *depend on* r, d *and* ε.

The proof relies again on the method of moments. Proposition 28 will be faster to prove than Proposition 24 since we do not need a sharp estimate of ρ_1. Let m be a positive integer. We argue as in (36)–(38). With the convention that $f_{2m+1} = f_1$, we get

$$\begin{aligned}
\mathbb{E}\|R_k^{(\ell)}\|^{2m} &\leq \mathbb{E}\mathrm{tr}\Big\{\Big(R_k^{(\ell)} R_k^{(\ell)^*}\Big)^m\Big\} \\
&= \sum_{(f_1,\dots,f_{2m}) \in E^{2m}} \mathbb{E}\mathrm{tr} \prod_{j=1}^m (R_k^{(\ell)})_{f_{2j-1},f_{2j}} (R_k^{(\ell)^*})_{f_{2j},f_{2j+1}} \\
&\leq \sum_{\gamma \in \widehat{W}_{\ell,m}} |\widehat{w}(\gamma)| \, \mathrm{tr}|a(\gamma)|,
\end{aligned}$$

where $a(\gamma)$ is as in (38), $\widehat{W}_{\ell,m}$ is the set of $\gamma = (\gamma_1, \dots, \gamma_{2m})$ such that for any $1 \leq j \leq m$, $\gamma_j = (\gamma_{j,1}, \dots, \gamma_{j,\ell+1}) \in F_k^{\ell+1}$, $\gamma_{j,t} = (x_{j,t}, i_{j,t})$, γ with the boundary condition (37), and we have set

$$\widehat{w}(\gamma) = \mathbb{E} \prod_{t=1}^{k-1} (\underline{S}_{i_{j,t}})_{x_{j,t} x_{j,t+1}} \prod_{t=k+1}^{\ell} (S_{i_{j,t}})_{x_{j,t} x_{j,t+1}}.$$

Using that $\max_i \|a_i\| \leq \varepsilon^{-1}$, we have $\mathrm{tr}|a(\gamma)| \leq r\varepsilon^{-\ell m}$ and thus,

$$\mathbb{E}\|R_k^{(\ell)}\|^{2m} \leq c^{2\ell m} \sum_{\gamma \in \widehat{W}_{\ell,m}} |\widehat{w}(\gamma)|. \tag{52}$$

To evaluate (52), we associate to each $\gamma \in \widehat{W}_{\ell,m}$, the graph $\widehat{G}_\gamma$ of visited vertices and colored edges that appear in the expression $\widehat{w}(\gamma)$. More precisely, for each j, we set $\gamma'_j = (\gamma_{j,1}, \dots, \gamma_{j,k}) \in F^k$ and $\gamma''_j = (\gamma_{j,k+1}, \dots, \gamma_{j,\ell+1}) \in F^{\ell+1-k}$. Then, with the notation in Definition 18, the vertex set of $\widehat{G}_\gamma$ is $V_\gamma = \bigcup_j V_{\gamma'_j} \cup V_{\gamma''_j}$ and the edge set is $\widehat{E}_\gamma = \bigcup_j E_{\gamma'_j} \cup E_{\gamma''_j}$. (For example, the black edge in Figure 5 is not part of $\widehat{E}_\gamma$.) The graph $\widehat{G}_\gamma$ may not be connected. However, due to the constraint on γ, it cannot have more vertices than edges. More precisely, let $\widehat{G}_{\gamma_j}$ denote the colored graph with vertex and edge sets $V_{\gamma'_j} \cup V_{\gamma''_j}$ and $E_{\gamma'_j} \cup E_{\gamma''_j}$. By the assumption that $\gamma_j \in F_k^{\ell+1}$, it follows that either $\widehat{G}_{\gamma_j}$ is a connected graph with a cycle or it has two connected components that both contain a cycle. Notably, since $\widehat{G}_\gamma$ is the union of these graphs, any connected component of $\widehat{G}_\gamma$ has a cycle, and it implies that

$$|V_\gamma| \leq |\widehat{E}_\gamma|. \tag{53}$$

Recall the definition of a canonical path above Lemma 25. The following lemma bound the number of canonical paths in $\widehat{W}_{\ell,m}$.

LEMMA 29. *Let $\widehat{\mathcal{W}}_{\ell,m}(v,e)$ be the subset of canonical paths in $\widehat{W}_{\ell,m}$ with $|V_\gamma| = v$ and $|\widehat{E}_\gamma| = e$. We have*

$$|\widehat{\mathcal{W}}_{\ell,m}(v,e)| \leq (2d\ell m)^{12m\chi+20m},$$

with $\chi = e - v + 1 \geq 1$.

Proof. The proof is identical to the proof of Lemma 25 up to the minor modification that for each j, γ'_j and γ''_j are tangle-free and non-backtracking (instead of simply γ_j). We use notation of Lemma 25: For $1 \leq j \leq 2m$, $1 \leq t \leq \ell$, $t \neq k$, we denote the visited edges by $e_{j,t} = (x_{j,t}, i_{j,t}, x_{j,t+1})$ and $[e_{j,t}] = [x_{j,t}, i_{j,t}, x_{j,t+1}] \in \widehat{E}_\gamma$. The graph $G_{(j,t)}$ is the graph spanned by the edges $\{[e_{j',t'}] : (j',t') \preceq (j,t), t' \neq k\}$, and $T_{(j,t)}$ is its spanning forest. For each j, we set $G_{(j,k)} = G_{(j,k-1)}$. The graphs $G_{(j,t)}$ are non-decreasing over time, and by definition $G_{(2m,\ell)} = \widehat{G}_\gamma$.

Now, for each j and γ'_j, γ''_j, the *merging times*, denoted by (j,σ') and (j,σ''), are the times such that γ'_j and γ''_j merge into a previous connected component. More precisely, if (j,t) with $1 \leq t \leq k-1$ (resp. $k+1 \leq t \leq \ell$) is the smallest time such that $x_{j,t+1}$ is a vertex of $G_{(j,1)^-}$ (resp. $G_{(j,k+1)^-}$) then $\sigma' = t$ (resp. $\sigma'' = t$). By convention, if $x_{j,1} \in G_{(j,1)^-}$, we set $\sigma' = 0$ (for example from (37), if j is odd, then $\sigma' = 0$), and we set $\sigma'' = k$ if $x_{j,k+1} \in G_{(j,k)^-}$. Similarly, we set $\sigma' = k$ (resp. $\sigma'' = \ell + 1$) if γ'_j does not interest $G_{(j,1)^-}$ (resp. γ''_j does not intersect $G_{(j,k)^-}$). *First times* and *important times* are defined as in Lemma 25.

We mark important times (j,t) by the vector $(i_{j,t}, x_{j,t+1}, x_{j,\tau}, i_{j,\tau})$, where (j,τ) is the next time that $[e_{j,\tau}]$ will not be a tree edge of the forest $T_{j,t}$ constructed so far. (By convention, if the path γ'_j or γ''_j remains on the forest, we set $\tau = k$ or $\tau = \ell + 1$.) For $t = 1$ and $t = k + 1$, we also add the *starting mark* $(x_{j,1}, \sigma', x_{j,\tau}, i_{j,\tau})$ and $(x_{j,k+1}, \sigma'', x_{j,\tau}, i_{j,\tau})$, where σ' and σ'' are the merging times and $(j,\tau) \geq (j,\sigma')$ or $(j,\tau) \geq (j,\sigma'')$ is, as above, the next time that $[e_{j,\tau}]$ will not be a tree edge of the forest constructed so far. As in Lemma 25, it gives rise to a first encoding $\widehat{\mathcal{W}}_{\ell,m}(v,e)$.

It can be improved by using that γ'_j and γ''_j are tangle-free. For each $1 \leq j \leq 2m$ and both for γ'_j and γ''_j, we define *short cycling*, *long cycling* and *superfluous* times as in Lemma 25 and we modify the mark of the short cycling time (j,t_s) as $(i_{j,t_s}, x_{j,t_s+1}, x_{j,t_1}, t_2, x_{j,\tau}, i_{j,\tau})$, where (j,t_1) is the closing time of the cycle, (j,t_2) is the exit time of the cycle, and $(j,\tau) \succeq (j,t_2)$ is the next time that $[e_{j,\tau}]$ will not be a tree edge of the forest constructed so far. For γ'_j (resp. γ''_j), important times (j,t) with $1 \leq t < t_s$ or $\tau \leq t \leq k-1$ (resp. $k+1 \leq t < t_s$ or $\tau \leq t \leq \ell$) are called long cycling times; they receive the usual mark $(i_{j,t}, x_{j,t+1}, x_{j,\tau}, i_{j,\tau})$. The other important times are called superfluous. By convention, if there is no short cycling time, the last important time is defined as the short cycling time. As argued in Lemma 25, there are at most $\chi - 1$ long cycling times for γ'_j and γ''_j.

This is the second encoding: we can reconstruct uniquely γ from the starting marks, the positions of the long cycling and the short cycling times and their marks. For each j, there are two starting marks, at most two short cycling times, and $2(\chi-1)$ long cycling times. There are at most $(\ell+1)^{4m\chi}$ ways to position them. There are at most d^2v^2 different possible marks for a long cycling time and $d^2v^3(\ell+1)$ possible marks for a short cycling time. Finally, there are $dv^2(\ell+1)$ possibilities for a starting mark. We deduce that

$$|\mathcal{W}_{\ell,m}(v,e)| \leq (\ell+1)^{4m\chi}(dv^2(\ell+1))^{4m}(d^2v^2)^{4m(\chi-1)}(d^2v^3(\ell+1))^{4m}.$$

Using $v \leq 2\ell m$ and $\ell+1 \leq 2\ell$, we obtain the claimed bound. □

We are ready to prove Proposition 28.

Proof of Proposition 28. For $n \geq 3$, we define

$$m = \left\lfloor \frac{\log n}{25\log(\log n)} \right\rfloor. \tag{54}$$

For this choice of m, $\ell m = o(\log n)^2$. From Markov inequality and (52), it suffices to prove that for some constants $c, c_1 > 0$,

$$S = \sum_{\gamma \in \widehat{W}_{\ell,m}} |\widehat{w}(\gamma)| \leq (c\ell m)^{32m} c_1^{2\ell m}. \tag{55}$$

From (53), $|V_\gamma| \leq |\widehat{E}_\gamma| \leq 2\ell m$ and

$$S \leq \sum_{v=1}^{2\ell m} \sum_{e=v}^{\infty} |\widehat{\mathcal{W}}_{\ell,m}(v,e)| \max_{\gamma \in \mathcal{W}_{\ell,m}(v,e)} (|\widehat{w}(\gamma)| N(\gamma)),$$

where $N(\gamma)$ is the number of γ' in $\widehat{W}_{\ell,m}$ such that $\gamma' \sim \gamma$. If $\gamma \in \mathcal{W}_{\ell,m}(v,e)$, the following trivial bound holds:

$$N(\gamma) \leq n^v d^e.$$

(Indeed, n^v bounds the possible choices for the vertices in V_γ and d^e the possible choices for the colors of the edges in $\widehat{E}_\gamma$.) Moreover, from Proposition 22 (bounding the number of inconsistent edges by e), if $\gamma \in \mathcal{W}_{\ell,m}(v,e)$, then

$$|\widehat{w}(\gamma)| \leq c\left(\frac{9}{n}\right)^e.$$

Using also Lemma 29, we find

$$S \le c \sum_{v=1}^{2\ell m} \sum_{e=v}^{\infty} (2d\ell m)^{12m(e-v)+32m} n^v d^e \left(\frac{9}{n}\right)^e,$$
$$= c(2d\ell m)^{32m} \sum_{v=1}^{2\ell m} (9d)^v \sum_{t=0}^{\infty} \left(\frac{9d(2d\ell m)^{12m}}{n}\right)^t.$$

For our choice of m, the geometric series is convergent. It follows that for some new constant $c, c' > 0$,

$$S \le c(2d\ell m)^{32m} \sum_{v=1}^{2\ell m} (9d)^v \le c'(2d\ell m)^{32m}(9d)^{2\ell m}.$$

This concludes the proof of (55) with $c_1 = 9d$. (A finer analysis as done in Proposition 24 leads to $c_1 = d - 1 + o(1)$.) □

4.5. *Proof of Theorem* 17. Let $0 < \varepsilon < 1$. For a given collection of weights $a = (a_i) \in M_r(\mathbb{C})^d$, we denote by $B(a)$ the corresponding non-backtracking operator and by $\mathcal{E}_\varepsilon(a)$ the event that $\rho(B(a)_{|K_0}) > \rho(B_\star(a)) + \varepsilon$. It is sufficient to prove that for some $\beta > 0$,

$$\mathbb{P}\left(\bigcup_{a \in \mathcal{S}_\varepsilon^d} \mathcal{E}_\varepsilon(a)\right) = O(n^{-\beta}), \tag{56}$$

where $\mathcal{B} \subset M_r(\mathbb{C})$ is the unit ball for the operator norm $\|\cdot\|$ and

$$\mathcal{S}_\varepsilon = \{b \in M_r(\mathbb{C}) : b \in \varepsilon^{-1}\mathcal{B}, b^{-1} \in \varepsilon^{-1}\mathcal{B}\}.$$

We use a net argument on $\mathcal{S}_\varepsilon^d$. Due to the lack of uniform continuity of spectral radii, we perform the net argument with operator norms. To this end, we fix an integer valued sequence $\ell(n) \sim (\log n)/\kappa$ for some $\kappa > 1$ satisfying

$$\kappa > \log\left((d-1)^4 \vee \left(\frac{4\rho_1}{\varepsilon}\right)\right),$$

where ρ_1 is as in Proposition 28. In order to lighten the notation, we will omit the dependence in n of $\ell(n)$ whenever appropriate. We denote by $\mathcal{E}'_\varepsilon(a)$ the event

$$\sup_{g \in K_0, \|g\|_2 = 1} \|B^\ell(a) g\|_2 > (\rho(B_\star(a)) + \varepsilon)^\ell.$$

From (27), the inclusion $\mathcal{E}_\varepsilon(a) \subset \mathcal{E}'_\varepsilon(a)$ holds. Moreover, from (12), $\|B(a)\| \le (d-1)\|a\|$, where

$$\|a\| = \sum_{i=1}^d \|a_i\|.$$

We deduce that the map $a \mapsto B^\ell(a)$ satisfies a deviation inequality

$$\begin{aligned}\|B^\ell(a) - B^\ell(a')\| &\leq \ell \max(\|B(a)\|, \|B(a')\|)^{\ell-1}\|B(a-a')\| \\ &\leq \ell(d-1)^\ell \max(\|a\|, \|a'\|)^{\ell-1}\|a-a'\|.\end{aligned} \tag{57}$$

We now build our net of $\mathcal{S}^d_\varepsilon$. First, since all matrix norms are equivalent and $M_r(\mathbb{C}) \simeq \mathbb{R}^{2r^2}$, we can find a subset $N_\delta \subset \varepsilon^{-1}\mathcal{B}$ of cardinality at most $(c/(\varepsilon\delta))^{2r^2}$ such that for any $b \in \varepsilon^{-1}\mathcal{B}$, there exists $b_0 \in N_\delta$ with $\|b - b_0\| \leq \delta$. (The constant c depends on r.) Note that

$$\|b_0^{-1}\| \leq \|b^{-1}\| + \|b_0^{-1} - b^{-1}\| \leq \|b^{-1}\| + \|b^{-1}\|\|b_0^{-1}\|\|b_0 - b\|.$$

Hence, if $\|b^{-1}\| \leq \varepsilon^{-1}$ and $\delta < \varepsilon/2$, we find

$$\|b_0^{-1}\| \leq \frac{\|b^{-1}\|}{1-\delta/\varepsilon} \leq \frac{2}{\varepsilon}.$$

We deduce that, if $\delta < \varepsilon/2$, there exists $N'_\delta \subset \mathcal{S}_{\varepsilon/2} \cap N_\delta$ such for any $b \in \mathcal{S}_\varepsilon$, there exists $b_0 \in N'_\delta$ with $\|b - b_0\| \leq \delta$. Consequently, if $\delta < \varepsilon d/2$, there exists a subset $N''_\delta = (N'_{\delta/d})^d \subset \mathcal{S}^d_{\varepsilon/2}$ of cardinal number at most $(cd/\varepsilon\delta)^{2r^2 d}$ such that for any $a \in \mathcal{S}^d_\varepsilon$, there exists $a_0 \in N''_\delta$ with $\|a - a_0\| \leq \delta$. Besides, from Lemma 14, for all δ small enough,

$$|\rho(B_\star(a)) - \rho(B_\star(a_0))| \leq \frac{\varepsilon}{3} \tag{58}$$

and, from (57), for some new constant $c > 0$,

$$\|B^\ell(a) - B^\ell(a_0)\| \leq \ell(d-1)\Big(\frac{d-1}{\varepsilon}\Big)^{\ell-1}\delta \leq c^\ell\delta.$$

If $\delta = (\varepsilon/3c)^\ell$ and if $\mathcal{E}_{\varepsilon/3}(a_0)$ does not hold, we deduce, for n large enough,

$$\begin{aligned}\sup_{g\in K_0, \|g\|_2=1}\|B^\ell(a)g\|_2 &\leq \sup_{g\in K_0, \|g\|_2=1}\|B^\ell(a_0)g\|_2 + \|B^\ell(a) - B^\ell(a_0)\| \\ &\leq \Big(\rho(B_\star(a_0)) + \frac{\varepsilon}{3}\Big)^\ell + \Big(\frac{\varepsilon}{3}\Big)^\ell \\ &\leq \Big(\rho(B_\star(a_0)) + \frac{2\varepsilon}{3}\Big)^\ell.\end{aligned}$$

Using (58), we find that, for our choice of δ and n large enough,

$$\bigcup_{a\in\mathcal{S}^d_\varepsilon}\mathcal{E}_\varepsilon(a) \subset \bigcup_{a\in\mathcal{S}^d_\varepsilon}\mathcal{E}'_\varepsilon(a) \subset \bigcup_{a\in N''_\delta}\mathcal{E}'_{\frac{\varepsilon}{3}}(a)$$

and, for some $c_1 > 0$ (depending on ε, r and d),

$$|N''_\delta| \leq c_1^\ell. \tag{59}$$

We may now use the union bound to obtain an estimate of (56). If Ω_0 is the event that G^σ is ℓ-tangle free, we find, for n large enough,

$$\mathbb{P}\Bigg(\bigcup_{a\in\mathcal{S}^d}\mathcal{E}_\varepsilon(a)\Bigg) \le \sum_{a\in N''_\delta}\mathbb{P}\Big(\mathcal{E}'_{\frac{\varepsilon}{3}}(a)\cap\Omega_0\Big)+\mathbb{P}(\Omega_0^c)$$

$$\le \sum_{a\in N''_\delta}\mathbb{P}\bigg(J(a)\ge\Big(\rho(B_\star(a))+\frac{\varepsilon}{3}\Big)^\ell\bigg)+O\bigg(\frac{\ell^3(d-1)^{4\ell}}{n}\bigg),$$

where at the second line, we have used Lemmas 20 and 23 and have set

$$J(a)=\|\underline{B}^{(\ell)}(a)\|+\frac{1}{n}\sum_{k=1}^{\ell}\|R_k^{(\ell)}(a)\|.$$

For our choice of ℓ, we note that $\ell^3(d-1)^{4\ell}/n=O(n^{-\beta})$ for some $\beta>0$. On the other end, by Propositions 24–28 applied to $\varepsilon'=\varepsilon/4$, for any $a\in\mathcal{S}_{\varepsilon'}$, with probability at least $1-c\ell\exp(-\ell\log n/(c\log\log n))$, we have

$$J(a)\le(\log n)^{20}\Big(\rho(B_\star(a))+\frac{\varepsilon}{4}\Big)^\ell$$

$$+\frac{1}{n}\sum_{k=1}^{\ell}(\log n)^{40}\rho_1^\ell\le(\log n)^c\Big(\rho(B_\star(a))+\frac{\varepsilon}{4}\Big)^\ell,$$

since $\ell=O(\log n)$ and (for n large enough) $\rho_1^\ell\le n(\varepsilon/4)^\ell$ thanks to our choice of ℓ. Finally, since $(\log n)^{c/\ell}=1+O(\log\log n/\log n)$, it follows that the event $\{J(a)\ge(\rho(B_\star(a))+\frac{\varepsilon}{3})^\ell\}$ holds with probability most

$$c\ell\exp(-\ell\log n/(c\log\log n)).$$

Using (59), we obtain

$$\mathbb{P}\Bigg(\bigcup_{a\in\mathcal{S}^d}\mathcal{E}_\varepsilon(a)\Bigg)=O\Big(\ell e^{-\frac{\ell\log n}{c\log\log n}}c_1^\ell+n^{-\beta}\Big)=O\big(n^{-\beta}\big).$$

The bound (56) follows. □

5. Proof of Theorem 4

We start with the inclusion (9) with $A^{(2)}$ in place of A. Note that $\ell^2(X^2)$ can be decomposed as the direct sum $\ell^2(X^2)=\ell^2(X^2_=)\oplus\ell^2(X^2_{\neq})$, where $X^2_= = \{(x,x):x\in X\}$ and $X^2_{\neq}=\{(x,y):x\neq y\in X\}$. Moreover, $A_{|\ell^2(X^2_=)}$ can be identified with A. It follows that the spectrum of $A^{(2)}$ contains the spectrum of A, and thus (9) holds also for $A^{(2)}$ thanks to Section 2.

We turn to the inclusion (10) with $A^{(2)}$ in place of A. Recall that the vector space V is spanned by I and J defined by (8). We set

$$K_0^{(2)}=\mathbb{C}^r\otimes V^\perp\otimes\mathbb{C}^d.$$

Arguing exactly as in the proof of Theorem 2, Theorem 4 is a consequence of the following statement on the non-backtracking operators

$$B = \sum_{j \neq i^*} a_j \otimes S_i \otimes S_i \otimes E_{ij}.$$

THEOREM 30. *Theorem 17 holds with A replaced by $A^{(2)}$, defined by (7), and K_0 is replaced by $K_0^{(2)}$.*

The proof of Theorem 30 follows essentially from the proof of Theorem 17. We now explain how to adapt the above argument. We shall follow the same steps and only highlight the differences.

The first observation is that part of Theorem 30 is already contained in Theorem 17. More precisely, as already pointed out, we have the direct sum $\ell^2(X^2) = \ell^2(X^2_=) \oplus \ell^2(X^2_{\neq})$. Then it is immediate to check that for any permutation operator S on $\ell^2(X)$, $S \otimes S$ decomposes orthogonally on $\ell^2(X^2_=) \oplus \ell^2(X^2_{\neq})$. Hence, B decomposes orthogonally on $(\mathbb{C}^r \otimes \ell^2(X^2_=) \otimes \mathbb{C}^d) \oplus (\mathbb{C}^r \otimes \ell^2(X^2_{\neq}) \otimes \mathbb{C}^d)$. Also, since $J \in \ell^2(X^2_{\neq})$ and $I \in \ell^2(X^2_=)$, the operator $B_{|K_0^{(2)}}$ decomposes orthogonally on $K_0^I \oplus K_0^J$ with

$$K_0^I = \mathbb{C}^r \otimes (I^\perp \cap \ell^2(X^2_=)) \otimes \mathbb{C}^d \quad \text{and} \quad K_0^J = \mathbb{C}^r \otimes (J^\perp \cap \ell^2(X^2_{\neq})) \otimes \mathbb{C}^d.$$

Finally, $B_{|K_0^I}$ can be identified with $B'_{|K_0}$, where $B' = \sum_{j \neq i^*} a_j \otimes S_i \otimes E_{ij}$. The spectral radius of $B'_{|K_0}$ can be bounded using Theorem 17. As a byproduct, it remains to prove Theorem 30 with $K_0^{(2)}$ replaced by K_0^J.

5.1. *Path decomposition.* We follow Section 4.2 and use the same notation. We set $X = [n]$ and let $A^{(2)}$ be as in (7). We now denote by $\boldsymbol{B}$ the non-backtracking matrix of $A^{(2)}$ restricted to $\mathbb{C}^r \otimes \ell^2(X^2_{\neq}) \otimes \mathbb{C}^d$. Our goal is to derive the analog of Lemma 20 for $\rho(\boldsymbol{B}_{|K_0^J})$. We define $\boldsymbol{E} = X^2_{\neq} \times [d]$. We may write $\boldsymbol{B}$ as a matrix-valued matrix on $\boldsymbol{E}$: for $e, f \in \boldsymbol{E}$, $e = (x, i)$, $f = (y, j)$, $x = (x^-, x^+)$, and $y = (y^-, y^+)$,

$$\boldsymbol{B}_{ef} = a_j \mathbf{1}(\sigma_i(x^-) = y^-) \mathbf{1}(\sigma_i(x^+) = y^+) \mathbf{1}(j \neq i^*) = a_j (S_i \otimes S_i)_{xy} \mathbf{1}(j \neq i^*).$$

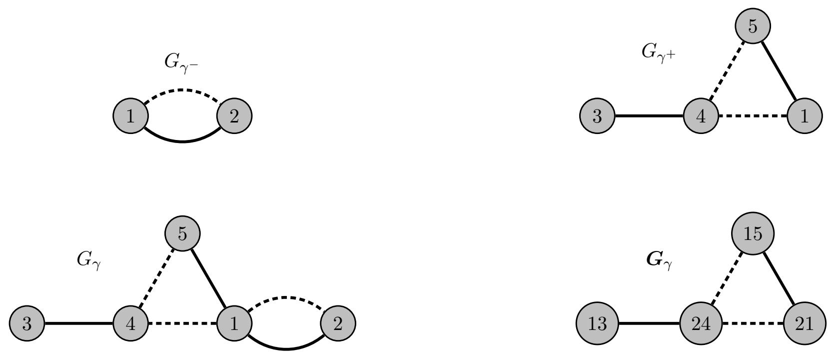
The next definitions extend Definitions 18–19. We revisit Definition 6, where we replace X by $X^2_{\neq}$, and σ_i by $\sigma_i \otimes \sigma_i$, $i \in [d]$. We may define a colored edge $[x, i, y]$ with $x, y \in X^2_{\neq}$, $i \in [d]$.

Definition 31. Let $\gamma = (\gamma_1, \dots, \gamma_k)$ in $\boldsymbol{E}^k$, $\gamma_t = (x_t, i_t)$, $x_t = (x_t^-, x_t^+)$.

- We set $\gamma^\pm = (\gamma_1^\pm, \dots, \gamma_k^\pm)$ with $\gamma_t^\pm = (x_t^\pm, i_t) \in E$.
- The *weight* of γ is $a(\gamma) = a(\gamma^\pm) = \prod_{t=2}^k a_{i_t}$.
- We set $\boldsymbol{V}_\gamma = \{x_t : 1 \leq t \leq k\}$, $\boldsymbol{E}_\gamma = \{[x_t, i_t, x_{t+1}] : 1 \leq t \leq k\}$, $V_\gamma = V_{\gamma^-} \cup V_{\gamma^+} = V_{(\gamma^-, \gamma^+)}$ and $E_\gamma = E_{\gamma^-} \cup E_{\gamma^+} = E_{(\gamma^-, \gamma^+)}$. We define the colored graphs $G_\gamma = (V_\gamma, E_\gamma)$ and $\boldsymbol{G}_\gamma = (\boldsymbol{V}_\gamma, \boldsymbol{E}_\gamma)$; see Figure 10.

– If e, f are in $\boldsymbol{E}$, we define $\boldsymbol{\Gamma}^k_{ef}$ as the subset of γ in $\boldsymbol{E}^k$, such that $(\gamma^-, \gamma^+) \in \Gamma^k_{e^-f^-} \times \Gamma^k_{e^+f^+}$. The sets $\boldsymbol{\Gamma}^k$, $\boldsymbol{F}^k$, $\boldsymbol{F}^{\ell+1}_k$ are defined in the same way from the sets Γ^k, F^k, $F^{\ell+1}_k$. For example, $\boldsymbol{F}^k$ is the set γ in $\boldsymbol{\Gamma}^k$ such that both (γ^-, γ^+) are tangle-free.

Note that according to this definition, if $\gamma \in \boldsymbol{F}^k$, then $\boldsymbol{G}_\gamma$ is necessarily tangle-free, while G_γ is not necessarily. See Figure 10 for an example.



$$\gamma = ((1,3),1)((2,4),2)((1,5),1)((2,1),2)((1,4),1)$$

Figure 10. A path $\gamma \in \boldsymbol{E}^5$ and its associated graphs $G_{\gamma^\pm}$, G_γ and $\boldsymbol{G}_\gamma$ (whose vertices have been written x^-x^+ instead of (x^-, x^+)), the involution i^* is the identity.

For e, f in $\boldsymbol{E}$, we find that

$$(\boldsymbol{B}^\ell)_{ef} = \sum_{\gamma \in \boldsymbol{\Gamma}^{\ell+1}_{ef}} a(\gamma) \prod_{t=1}^{\ell} (S_{i_t} \otimes S_{i_t})_{x_t x_{t+1}}.$$

The orthogonal projection of $S_i \otimes S_i$ on $J^\perp \cap \ell^2(X^2_{\neq})$ is given by

$$(\underline{S_i \otimes S_i}) = (S_i \otimes S_i) - P_J.$$

where, for any $g \in \ell^2(X^2_{\neq})$, $P_J(g) = J\langle J, g\rangle/(n(n-1))$. Alternatively, in matrix form, for any x, y in $X^2_{\neq}$,

$$(\underline{S_i \otimes S_i})_{xy} = (S_i \otimes S_i)_{xy} - \frac{1}{n(n-1)}. \tag{60}$$

Now, recall the definition of the colored graph G^σ with vertex set $[n]$ in Definition 6. Obviously, if G^σ is ℓ-tangle-free and $0 \le k \le 2\ell$, then

$$\boldsymbol{B}^k = \boldsymbol{B}^{(k)}, \tag{61}$$

where

$$(\boldsymbol{B}^{(k)})_{ef} = \sum_{\gamma \in \boldsymbol{F}^{k+1}_{ef}} a(\gamma) \prod_{t=1}^{k} (S_{i_t} \otimes S_{i_t})_{x_t x_{t+1}}.$$

We define similarly the matrix $\underline{B}^{(k)}$ by

(62) $$(\underline{\boldsymbol{B}}^{(k)})_{ef} = \sum_{\gamma \in \boldsymbol{F}_{ef}^{k+1}} a(\gamma) \prod_{t=1}^{k} (\underline{S_{i_t} \otimes S_{i_t}})_{x_t x_{t+1}}.$$

We may now use the telescopic sum decomposition performed in Section 4.2. We denote by $\overline{\boldsymbol{B}}$ the matrix on $\mathbb{C}^r \otimes \mathbb{C}^{\boldsymbol{E}}$ defined by

$$\overline{\boldsymbol{B}} = \sum_{j \neq i^*} a_j \otimes P_J \otimes E_{ij}.$$

We also set for all $e, f \in \boldsymbol{E}$,

(63)
$$(\boldsymbol{R}_k^{(\ell)})_{ef} = \sum_{\gamma \in \boldsymbol{F}_{k,ef}^{\ell+1} \setminus \boldsymbol{F}_{ef}^{\ell+1}} a(\gamma) \left(\prod_{t=1}^{k-1} (\underline{S_{i_t} \otimes S_{i_t}})_{x_t x_{t+1}} \right) \left(\prod_{t=k+1}^{\ell} (S_{i_t} \otimes S_{i_t})_{x_t x_{t+1}} \right).$$

Arguing as in Section 4.2, we have

$$\boldsymbol{B}^{(\ell)} = \underline{\boldsymbol{B}}^{(\ell)} + \frac{1}{n(n-1)} \sum_{k=1}^{\ell} \underline{\boldsymbol{B}}^{(k-1)} \overline{\boldsymbol{B}} \boldsymbol{B}^{(\ell-k)} - \frac{1}{n(n-1)} \sum_{k=1}^{\ell} \boldsymbol{R}_k^{(\ell)}.$$

Now, if G^σ is ℓ-tangle free, then from (61), $\overline{\boldsymbol{B}} \boldsymbol{B}^{(\ell-k)} = \overline{\boldsymbol{B}} \boldsymbol{B}^{\ell-k}$. Since the kernel of $\overline{\boldsymbol{B}}$ contains K_0^J and $\boldsymbol{B}^{\ell-k} K_0^J \subset K_0^J$, we find that $\overline{\boldsymbol{B}} \boldsymbol{B}^{\ell-k} = 0$ on K_0^J. So finally, if G^σ is ℓ-tangle free, then for any $g \in K_0^J$,

$$\boldsymbol{B}^{(\ell)} g = \underline{\boldsymbol{B}}^{(\ell)} g - \frac{1}{n(n-1)} \sum_{k=1}^{\ell} \boldsymbol{R}_k^{(\ell)} g.$$

We get the following lemma.

LEMMA 32. *Let $\ell \geq 1$ be an integer, and let $A^{(2)}$ as in (7) be such that G^σ is ℓ-tangle free. Then,*

$$\rho(\boldsymbol{B}_{|K_0^J}) \leq \left(\|\underline{\boldsymbol{B}}^{(\ell)}\| + \frac{1}{n(n-1)} \sum_{k=1}^{\ell} \|\boldsymbol{R}_k^{(\ell)}\| \right)^{1/\ell}.$$

5.2. *Novel estimate on random permutations.* In this subsection, we prove the analog of Proposition 22 for the tensors $S_i \otimes S_i$ for the symmetric random permutations. To this end, we need to adapt the proof of this proposition in [9].

Consider a sequence of colored edges $(f_1, \ldots, f_\tau)$ with $f_t = [x_t, i_t, y_t]$ and $x_t = (x_t^-, x_t^+), y_t = (y_t^-, y_t^+) \in X_{\neq}$. We set $f_t^\pm = (x_t^\pm, i_t, y_t^\pm)$. We say an edge $[x, i, y] \in \{f_t : 1 \leq t \leq \tau\}$ is *consistent* if $[x^-, i, y^-]$ *and* $[x^+, i, y^+]$ are consistent in the colored graph spanned by the edge $\{f_t^-, f_t^+ : 1 \leq t \leq \tau\}$ (where the definition of a consistent edge is given in Definition 21). It is inconsistent otherwise. Its *multiplicity* is the pair of multiplicities (m^-, m^+) of

$[x^-, i, y^-]$ and $[x^+, i, y^+]$ in the sequence $(f_1^+, \dots, f_\tau^+, f_1^-, \dots, f_\tau^-)$. (That is, $m^\pm = \sum_t \mathbf{1}(f_t^- = [x^\pm, i, y^\pm]) + \mathbf{1}(f_t^+ = [x^\pm, i, y^\pm])$.)

PROPOSITION 33. *For symmetric random permutations, there exists a constant $c > 0$ such that for any sequence $(f_1, \dots, f_\tau)$, with $f_t = [x_t, i_t, y_t]$, $x_t, y_t \in X_{\neq}$, $4\tau \le n^{1/3}$ and any $\tau_0 \le \tau$, we have*

$$\left| \mathbb{E} \prod_{t=1}^{\tau_0} (\underline{S_{i_t} \otimes S_{i_t}})_{x_t y_t} \prod_{t=\tau_0+1}^{\tau} (S_{i_t} \otimes S_{i_t})_{x_t y_t} \right| \le c\, 9^b \left(\frac{1}{n}\right)^e \left(\frac{6\tau}{\sqrt{n}}\right)^{e_1},$$

where $e = |\{f_t^+, f_t^- : 1 \le t \le \tau\}|$, b is the number of inconsistent edges of and e_1 is the number of $1 \le t \le \tau_0$ such that f_t is consistent and has multiplicity $(1,1)$.

We will use the Pochammer symbol, defined for non-negative integers a, b,

$$(a)_b = \prod_{p=0}^{b-1} (a - p).$$

(Recall the convention that a product over an empty set is 1.) We will use the following lemma, whose proof is postponed to Section 7.

LEMMA 34. *Let $z \ge 1$, $k \ge 1$ be an integer, $0 < p, q \le 1/4$ and N be a* $\mathrm{Bin}(k,p)$ *variable. If $8(1 - p - p/q)^2 \le 4zk^2\sqrt{q} \le 1$, we have*

$$\left| \mathbb{E} (-1)^N \prod_{n=0}^{2N-1} \left(\frac{1}{\sqrt{q}} - zn \right) \right| \le 8(3\sqrt{2z}kq^{1/4})^k.$$

Proof of Proposition 33. We adapt the proof of [9, Prop. 8]. Using the independence of the matrices S_i (up to the involution), it is enough to consider the case of a single permutation matrix. We set $S = S_i$ and $\sigma = \sigma_i$. The colored edge $[x, i, y]$ of $\{f_t : 1 \le t \le \tau\}$ is simply denoted by (x, y). Note that we may view (x, y) as an oriented edge from x to y as i is fixed. We treat the case of S uniformly sampled random permutation. (The case of uniform matching is similar, see final comment below.) We will repeatedly use that if $1 \le k \le a\sqrt{n}$, then for some $c = c(a) > 0$,

$$(n - k)^{-k} \le cn^{-k}.$$

We first assume that all edges are consistent. Let $X = \{x_t^\varepsilon : t \in [\tau], \varepsilon \in \{-,+\}\}$, $Y = \{y_t^\varepsilon : t \in [\tau], \varepsilon \in \{-,+\}\}$ and $\{g_1, \dots, g_e\} = \{(x_t^\varepsilon, y_t^\varepsilon) : t \in [\tau], \varepsilon \in \{-,+\}\}$ with $g_s = (u_s, v_s)$ be the distinct edges of $\{f_t^\varepsilon : t \in [\tau], \varepsilon \in \{-,+\}\}$. Let T be the set $1 \le t \le \tau_0$ such that (x_t, y_t) has multiplicity $(1,1)$. Let $T_2 \subset T$ be the subset of t in T such that $(\sigma(x_t^-), \sigma(x_t^+)) \in \{(y_t^-, y_t^+)\} \cup ([n] \backslash Y)^2$ and $(\sigma^{-1}(y_t^-), \sigma^{-1}(y_t^+)) \in \{(x_t^-, x_t^+)\} \cup ([n] \backslash X)^2$. In words, elements in T_2 are either matched perfectly $((\sigma(x_t^-), \sigma(x_t^+) = (y^+, y^-))$ or have their image and preimage outside of γ. We set $T_1 = T\backslash T_2$. By construction, if $t \in T_1$, then

$$\underline{(S \otimes S)}_{x_t y_t} = (-m)^{-1},$$

with $m = n(n-1)$. We may thus write

$$P = \prod_{t=1}^{\tau_0} \underline{(S \otimes S)}_{x_t y_t} \prod_{t=\tau_0+1}^{\tau} (S \otimes S)_{x_t y_t} = (-m)^{-|T_1|} \times P_2 \times P_3, \tag{64}$$

where

$$P_2 = \prod_{t \in T_2} \underline{(S \otimes S)}_{x_t y_t} \quad \text{and} \quad P_3 = \prod_{t \in [\tau_0] \backslash T_1 \cup T_2} \underline{(S \otimes S)}_{x_t y_t} \prod_{t=\tau_0+1}^{\tau} (S \otimes S)_{x_t y_t}.$$

Let $\mathcal{F}$ be the filtration generated by the variables T_2 and $(\sigma(x_t^{\pm}), \sigma^{-1}(y_t^{\pm}))$, $t \in [\tau]\backslash T_2$. By construction, the variable T_1 and P_3 are $\mathcal{F}$-measurable. If $\mathbb{E}_{\mathcal{F}}[\cdot]$ denotes the conditional expectation given $\mathcal{F}$, it follows that

$$|\mathbb{E}P| = \Big|\mathbb{E}\Big[(-m)^{-|T_1|} P_3 \mathbb{E}_{\mathcal{F}}[P_2]\Big]\Big| \le c\,\mathbb{E}\Big[n^{-2|T_1|} \cdot \mathbb{E}[|P_3||T_1] \cdot |\mathbb{E}_{\mathcal{F}}[P_2]|\Big]. \tag{65}$$

We start by estimating P_3 in (65). Consider the graph, say Γ, with vertex set $\{g_1, \dots, g_e\}$ and whose τ edges are $\{(x_t^-, y_t^-), (x_t^+, y_t^+)\}$. (It may have loops and multiple edges.) Let L be set of g_s, $1 \le s \le e$, such that $\sigma(u_s) \neq v_s$ and $g_s \neq (x_t^{\varepsilon}, y_t^{\varepsilon})$ for all $t \in T$, $\varepsilon \in \{-,+\}$. Let K be the set of edges of Γ adjacent to a vertex in L. We have

$$|P_3| \le m^{-|K|}.$$

(If the t-th edge is adjacent to an element in L, then $|\underline{(S \otimes S)}_{x_t y_t}| = m^{-1}$ and $(S \otimes S)_{x_t y_t} = 0$.) We claim that

$$|K| \ge \frac{2|L|}{3}.$$

Indeed, consider the subgraph spanned by the edges in K. This subgraph has vertex set $L' \supseteq L$. Consider a connected component of this graph, with vertex set L_0' and edge set K_0. Set $L_0 = L \cap L_0'$. It is sufficient to check that $|K_0| \ge 2|L_0|/3$. If $|L_0'| \ge 3$, then the claimed bound follows from $|K_0| \ge |L_0'| - 1$. Similarly, if $|L_0'| = 1$, we have $L_0' = L_0$ and $|K_0| \ge 1$. If $|L_0'| = 2$ and $|K_0| \ge 2$, the bound holds trivially. The last remaining case is $|L_0'| = 2$ and $|K_0| = 1$. This case follows from the claim $|L_0| = 1$. Indeed, since $|K_0| = 1$, if both vertices, say g_a, g_b of this connected component are in L, then by construction, there is a unique t_0 such that g_a or g_b are in $\{(x_{t_0}^{\varepsilon}, y_{t_0}^{\varepsilon}) : \varepsilon \in \{-,+\}\}$ and $\{g_a, g_b\} = \{(x_{t_0}^-, y_{t_0}^-), (x_{t_0}^+, y_{t_0}^+)\}$. It implies that (x_{t_0}, y_{t_0}) has multiplicity $(1,1)$. It contradicts the definition of L (which contains no g_t with multiplicity $(1,1)$). So finally, we have proven that

$$|P_3| \le m^{-2|L|/3}. \tag{66}$$

We now estimate the law of the random variable $|L|$. It follows from equation (65) that we have to estimate $\mathbb{P}(|L| = x | T_1)$. For $t \in [e]$, let $\mathcal{F}_t$ be the the filtration generated by the variables $\sigma(u_s), \sigma^{-1}(v_s), s \in [e] \backslash \{t\}$. We have

$$\mathbb{P}(S_{u_t v_t} = 1 | \mathcal{F}_t) \leq \frac{1}{\hat{n}}, \tag{67}$$

where we introduced the notation $\hat{n} = n - \tau + 1$. Hence, if $e_2 = e - 2e_1$, it follows that for any integer $0 \leq x \leq e_2$,

$$\mathbb{P}(|L| = x | T_1) \leq \binom{e_2}{x} (\hat{n})^{x - e_2} \leq (2\tau)^x (\hat{n})^{x - e_2}.$$

From (66) and using $4\tau \leq n^{1/3}$, we get for some $c > 0$,

$$\mathbb{E}[|P_3| | T_1] \leq \sum_{x=0}^{\infty} (2\tau)^x (\hat{n})^{x - e_2} m^{-\frac{2x}{3}} \leq (\hat{n})^{-e_2} \sum_{x=0}^{\infty} 2^{-x} \left(\frac{n^{4/3}}{m^{2/3}} \right)^x \leq c n^{-e_2}. \tag{68}$$

We now give an upper bound on $\mathbb{E}_{\mathcal{F}}[P_2]$ in (65). This is where Lemma 34 is used. Let τ_2 be the number of $t \in T_2$ such that $(\sigma(x_t^-), \sigma(x_t^+)) \neq (y_t^-, y_t^+)$. We have

$$\mathbb{E}_{\mathcal{F}} P_2 = \mathbb{E}_{\mathcal{F}} \Big(1 - \frac{1}{m} \Big)^{|T_2| - \tau_2} \Big(-\frac{1}{m} \Big)^{\tau_2}.$$

Let $\bar{n} = n - |Y| - \sum_{t \notin T_2} \mathbf{1}(\sigma(u_t) \notin Y)$. By a direct counting argument, the law of τ_2 given $\mathcal{F}$ is given, for $0 \leq x \leq |T_2|$, by

$$\mathbb{P}_{\mathcal{F}}(\tau_2 = x) = \frac{\binom{|T_2|}{x} (\bar{n})_{2x}}{Z} \quad \text{with} \quad Z = \sum_{x=0}^{|T_2|} \binom{|T_2|}{x} (\bar{n})_{2x}. \tag{69}$$

Indeed, we use the fact that a uniform law conditioned by an event remains uniform. In turn, the term $\binom{|T_2|}{x}$ accounts for the possible choices of $t \in T_2$ such that $(\sigma(x_t^-), \sigma(x_t^+)) \neq (y_t^-, y_t^+)$. Once these t have been chosen, we use that for all $t \in T_2$, $(\sigma(x_t^-), \sigma(x_t^+)) \in \{(y_t^-, y_t^+)\} \cup ([n] \backslash Y)^2$ and $(\sigma^{-1}(y_t^-), \sigma^{-1}(y_t^+)) \in \{(x_t^-, x_t^+)\} \cup ([n] \backslash X)^2$. There are $(\bar{n})_{2x}$ such choices. It is immediate to estimate Z. Indeed, since $\bar{n} \geq n - 4k$, $|T_2| \leq k$ and $k \ll \sqrt{n}$, we find, for some $c > 0$,

$$Z \geq \sum_{x=0}^{|T_2|} \binom{|T_2|}{x} (n - 6k)^{2x} = [(n - 6k)^2 + 1]^{|T_2|} \geq c n^{2|T_2|}.$$

Also, we deduce that

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}} P_2 &= \frac{1}{Z} \sum_{x=0}^{|T_2|} \binom{|T_2|}{x} (-1)^x \prod_{y=0}^{2x-1} (\bar{n} - y) \Big(1 - \frac{1}{m} \Big)^{|T_2| - x} \Big(\frac{1}{m} \Big)^x \\
&= \frac{1}{Z} \mathbb{E} (-1)^{\tau_2} \prod_{y=0}^{2\tau_2 - 1} (\bar{n} - y),
\end{aligned}$$

where N has distribution $\mathrm{Bin}(|T_2|, 1/m)$. By Lemma 34 applied to $z = 1$, $k = |T_2|$ (which is at most k), $p = 1/m$, $q = 1/\bar{n}^2$, we deduce that, with $\varepsilon = 3\sqrt{2}k/\sqrt{n}$, for some $c > 0$,

$$|\mathbb{E}_{\mathcal{F}} P_2| \leq c\left(\frac{\varepsilon}{n^2}\right)^{|T_2|}. \tag{70}$$

Since $|T_2| + |T_1| = e_1$ and $2e_1 + e_2 = e$, we obtain in (65) from (68) and (70) that, for some $c > 0$,

$$|\mathbb{E}P| \leq cn^{-e}\varepsilon^{e_1}\mathbb{E}\varepsilon^{-|T_1|}.$$

It thus remains to show that $\mathbb{E}\varepsilon^{-|T_1|}$ is of order 1. If $|T_1| = x$, then there are least $\lceil x/2\rceil$ distinct x_t^ε with $t \in [k], \varepsilon \in \{-,+\}$ such that $\sigma(x_t^\varepsilon) \in Y$. From (67), we find that

$$\mathbb{P}(|T_1| = x) \leq \binom{2k}{\lceil x/2\rceil}\binom{|Y|}{\lceil x/2\rceil}(\hat{n})^{-\lceil x/2\rceil} \leq c\left(\frac{2k}{\sqrt{n}}\right)^{2\lceil x/2\rceil} \leq c\left(\frac{2k}{\sqrt{n}}\right)^{x}.$$

We deduce

$$\mathbb{E}\varepsilon^{-|T_1|} \leq c\sum_{x=0}^{\infty}\varepsilon^{-x}\left(\frac{2k}{\sqrt{n}}\right)^{x} = c\sum_{x=0}^{\infty}(\sqrt{2}/3)^x.$$

The latter series is convergent, and it concludes the proof when all edges are consistent.

We now extend to the case of inconsistent edges. Let us say that $x \in \boldsymbol{X} = \{x_t : t \in [\tau]\}$ is inconsistent with degree $\delta \geq 2$ if there are $y_1, \ldots, y_\delta$ distinct elements of $\boldsymbol{Y} = \{y_t : t \in [\tau]\}$ such that for any $1 \leq s \leq \delta$, (x, y_s) is in $\{f_t : t \in [\tau]\}$. We define similarly the degree of an inconsistent vertex $\boldsymbol{Y}$. The (vertex) inconsistency of $\boldsymbol{f} = (f_1, \cdots, f_\tau)$ is defined as the sum of the degrees of inconsistent vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$. The inconsistency of $\boldsymbol{f}$, say $\hat{b}$, is at most $2b$. Assume that x is an inconsistent vertex of degree $\delta \geq 2$ and (x, y_1) and (x, y_2) are in $\{f_t : t \in [\tau]\}$. Observe that $(S \otimes S)_{xy_1}(S \otimes S)_{xy_2} = 0$, hence, for any integers $p_1, p_2 \geq 1$,

$$\begin{aligned}
&(\underline{S \otimes S})_{xy_1}^{p_1}(\underline{S \otimes S})_{xy_2}^{p_2} \\
&\quad = \left((S \otimes S)_{xy_1} - \frac{1}{m}\right)^{p_1}\left((S \otimes S)_{xy_2} - \frac{1}{m}\right)^{p_2} \\
&\quad = (\underline{S \otimes S})_{xy_1}^{p_1}\left(-\frac{1}{m}\right)^{p_2} + \left(-\frac{1}{m}\right)^{p_1}(\underline{S \otimes S})_{xy_2}^{p_2} - \left(-\frac{1}{m}\right)^{p_1+p_2}.
\end{aligned}$$

Similarly, if $q_1 \geq 1$, then

$$(\underline{S \otimes S})_{xy_1}^{p_1}(S \otimes S)_{xy_1}^{q_1}(\underline{S \otimes S})_{xy_2}^{p_2} = (\underline{S \otimes S})_{xy_1}^{p_1}(S \otimes S)_{xy_1}^{q_1}\left(-\frac{1}{m}\right)^{p_2}.$$

If p_i, q_i is the number of occurrences of $(\underline{S \otimes S})_{xy_i}$ and $(S \otimes S)_{xy_i}$ in the product (64), we thus have decomposed (64) into at most three terms of the same form as (64) up to a factor $(-m)^{-p'}$. Each of these terms is associated to a new sequence $\boldsymbol{f}'$ of colored edges that is a subsequence of $\boldsymbol{f}$ with $e' \leq e$ distinct elements and $e'_1 \geq e_1$ elements of multiplicity $(1,1)$. Moreover we have $p'/2 + e' \geq e$ and the inconsistency of $\boldsymbol{f}'$ is at most $\hat{b} - 1$. By repeating this for all inconsistent vertices, we may decompose (64) into at most than $3^{\hat{b}} \leq 9^b$ terms of the form (64) with $e' \leq e$, $e'_1 \geq e_1$, where all edges are consistent, multiplied by a factor $(-m)^{-p'}$ with $p'/2 + e' \geq e$. Applying the first part of the proposition to each term, the conclusion follows.

Case where σ is a uniform matching. The proof follows from the same line. The only noticeable difference is for the distribution of the random variable τ_2 in (69). We will apply Lemma 34 with $z = 2$ (see [9, Prop. 8]). □

5.3. *Trace method of Füredi and Komlós.* We adapt here the content of Section 4.4 to tensors of permutation matrices. We explain here how to perform this adaptation.

We fix a collection $(a_i), i \in [d]$, of matrices such that the symmetry condition (3) holds, and we assume $\max_i(\|a_i\| \vee \|a_i^{-1}\|^{-1}) \leq \varepsilon^{-1}$ for some $\varepsilon > 0$. Then, $B_\star$ is the corresponding non-backtracking operator in the free group. The constants may depend implicitly on r, d and ε. We have the following analogs of Propositions 24 and 28.

PROPOSITION 35. *Let $\varepsilon > 0$. If $1 \leq \ell \leq \log n$, then the event*

$$\|\underline{\boldsymbol{B}}^{(\ell)}\| \leq (\log n)^{50}(\rho(B_\star) + \varepsilon)^\ell$$

holds with the probability at least $1 - ce^{-\frac{\ell \log n}{c \log\log n}}$, where $c > 0$ depends on r, d and ε.

PROPOSITION 36. *For any $1 \leq k, \ell \leq \log n$, the event*

$$\|\boldsymbol{R}_k^{(\ell)}\| \leq (\log n)^{100}\rho_1^\ell$$

holds with the probability at least $1 - ce^{-\frac{\ell \log n}{c \log\log n}}$, where $c > 0$ and $\rho_1 > 0$ depend on r, d and ε.

We only explain the differences arising in the proof of Proposition 35, the case of Proposition 36 being identical. Let m be a positive integer. The computation leading to (38) gives

$$\mathbb{E}\|\underline{\boldsymbol{B}}^{(\ell)}\|^{2m} \leq \sum_{\gamma \in \boldsymbol{W}_{\ell,m}} |w(\gamma)| \operatorname{tr}|a(\gamma)|, \tag{71}$$

where $\boldsymbol{W}_{\ell,m}$ is the set of $\gamma = (\gamma_1, \ldots, \gamma_{2m})$ such that $\gamma_j = (\gamma_{j,1}, \ldots, \gamma_{j,\ell+1}) \in \boldsymbol{F}^{\ell+1}$, $\gamma_{j,t} = (x_{j,t}, i_{j,t})$ and for all $j = 1, \ldots, m$, the boundary condition $\gamma_{2j,1} = \gamma_{2j+1,1}$ and $\gamma_{2j-1,\ell+1} = \gamma_{2j,\ell+1}$, with the convention that $\gamma_{2m+1} = \gamma_1$. In (71), we have also set

$$w(\gamma) = \mathbb{E}\prod_{t=1}^{\ell}(\underline{S_{i_{j,t}} \otimes S_{i_{j,t}}})_{x_{j,t}x_{j,t+1}} \quad \text{and} \quad a(\gamma) = \prod_{j=1}^{2m} a(\gamma_j)^{\varepsilon_j},$$

and $a(\gamma_j)^{\varepsilon_j}$ is $a(\gamma_j)$ or $a(\gamma_j)^*$ depending on the parity of j.

Let $X = [n]$. Exactly as in the proof of Proposition 24, we define the isomorphism class $\gamma \sim \gamma'$ if there exist permutations $\sigma \in \mathcal{S}_{X^2_{\neq}}$ and $(\tau_x)_x \in (\mathcal{S}_d)^{X^2_{\neq}}$ such that, with $\gamma'_{j,t} = (x'_{j,t}, i'_{j,t})$, for all $1 \le j \le 2m$, $1 \le t \le \ell+1$, $x'_{j,t} = \sigma(x_{j,t})$, $i'_{j,t} = \tau_{x_{j,t}}(i_{j,t})$ and $(i'_{j,t})^* = \tau_{x_{j,t+1}}((i_{j,t})^*)$. For each $\gamma \in W_{\ell,m}$, we define γ_-, γ_+ and G_γ as in Definition 31. The vertex set of G_γ is $V_\gamma = \cup_j V_{\gamma_j}$, and its edge set is $E_\gamma = \cup_j E_{\gamma_j}$. We also define $\boldsymbol{V}_\gamma$, $\boldsymbol{E}_\gamma$ and $\boldsymbol{G}_\gamma$ as in Definition 31. The graph G_γ is the union of the graphs G_{γ^-} and G_{γ^+}. Since $G_{\gamma^\pm}$ is connected, G_γ has at most two connected components, and it follows that $|E_\gamma| - |V_\gamma| + 2 \ge 0$. We may also define a canonical element in each isomorphic class as in the proof of Proposition 24. The analog of Lemma 25 is the following.

LEMMA 37. *Let $\boldsymbol{\mathcal{W}}_{\ell,m}(v,e)$ be the subset of canonical paths with $|V_\gamma| = v$ and $|E_\gamma| = e$. We have*

$$|\boldsymbol{\mathcal{W}}_{\ell,m}(v,e)| \le (4d\ell m)^{12m\chi+20m},$$

with $\chi = e - v + 2 \ge 0$.

Proof. We repeat the proof of Lemma 25, where we replace γ by the sequence (γ^-, γ^+), it then essentially amounts to replace $2m$ by $4m$. (The merging time of γ_1^+ may be empty if G_{γ^-} and G_{γ^+} are disjoint.) □

There is also an analog of Lemma 26.

LEMMA 38. *Let $\rho = \rho(B_\star) + \varepsilon$. For any positive integer k_0, there exists a constant $c > 0$ depending on r, d and ε such that for any $\gamma \in \boldsymbol{\mathcal{W}}_{\ell,m}(v,e)$,*

$$\sum_{\gamma':\gamma'\sim\gamma} \operatorname{tr}|a(\gamma')| \le c^{m+\chi+e_1} n^v \rho_0^{2(\ell m - v)} \rho^{2\ell m}, \tag{72}$$

where $\chi = e - v + 2$, e_1 is the number of edges of $\boldsymbol{E}_\gamma$ multiplicity $(1,1)$. Moreover, in the above equation, we have set

$$\rho_0 = \max \left\| \prod_{s=1}^{k_0} a_{i_s} \right\|^{\frac{1}{k_0}},$$

where the maximum is over all non-backtracking sequences $(i_1, \ldots, i_{k_0})$, that is $i_{s+1} \ne i_{s^*}$. Moreover, for all k_0 large enough, we have $\rho_0 \le \rho$.

Proof. In order to adapt the proof of Lemma 26, we may consider the graph $\boldsymbol{G}_\gamma$ with vertex set $\boldsymbol{V}_\gamma$ and colored edges $\boldsymbol{E}_\gamma$. An issue is that edges visited once on this graph by γ are not necessarily edges of multiplicity $(1,1)$. For example, in Figure 10 all edges of $\boldsymbol{E}_\gamma$ are visited exactly once but none of them is of multiplicity $(1,1)$.

To circumvent this difficulty, we introduce a new graph. Consider the following equivalence class on $\boldsymbol{V}_\gamma$: $x \sim x'$ if there exists a sequence $y_0, \dots, y_t$ in $\boldsymbol{V}_\gamma$ such that $y_0 = x$, $y_t = x'$ and for all $s \in [t]$, $\{y^-_{s-1}, y^+_{s-1}\} \cap \{y^-_s, y^+_s\} \neq \emptyset$ (In words, we glue iteratively together elements of γ that share some common vertices of G_γ.) If $\widetilde{V_\gamma}$ is the set of equivalence classes for this equivalence relation, we may define the graph $\widetilde{G_\gamma} = (\widetilde{V_\gamma}, \widetilde{E_\gamma})$ as the quotient graph of $\boldsymbol{G}_\gamma$. More precisely, $\widetilde{E_\gamma}$ is obtained from $\boldsymbol{E}_\gamma$ by identifying two edges $[u, i, u']$ and $[v, i, v']$ if there exists a sequence $[x_0, i, y_0], \dots, [x_t, i, y_t]$ in $\boldsymbol{E}_\gamma$ such that $(x_0, i, y_0) = (u, i, u')$, $(x_t, i, y_t) = (v, i, v')$ and for all $s \in [t]$,

$$\{(x^-_{s-1}, i, y^-_{s-1}), (x^+_{s-1}, i, y^+_{s-1})\} \cap \{(x^-_s, i, y^-_s), (x^+_s, i, y^+_s)\} \neq \emptyset.$$

For example if $\gamma \in \boldsymbol{E}^5$ is the path of Figure 10, then $\widetilde{V_\gamma}$ has a single vertex with two colored loops attached. If $\widetilde{e} = |\widetilde{E_\gamma}|$, $\widetilde{v} = |\widetilde{V_\gamma}|$ and $\widetilde{\chi} = \widetilde{e} - \widetilde{v} + 1$, by iteration on the successive gluings of vertices of G_γ, we find easily that $\widetilde{\chi} \leq \chi$.

Similarly, we define $\widetilde{\gamma} = ((\widetilde{x}_{j,t}, i_{j,t}))_{j,t}$ from the original path $\gamma = ((x_{j,t}, i_{j,t}))$, $1 \leq j \leq 2m, 1 \leq t\ell + 1$. Then, by construction, the number of edges of multiplicity 1 in $\widetilde{\gamma}$, say $\widetilde{e}_1$, is at most e_1, the number of edges of multiplicity $(1,1)$ in γ. We may then simply repeat the proof of Lemma 26 with $\widetilde{\gamma}$ and $\widetilde{G}_\gamma$ in place of γ and G_γ. □

We finally give the analog of Lemma 27 for $w(\gamma)$ defined below (71).

LEMMA 39. *There exists a constant $c > 0$ such that for all $\gamma \in \boldsymbol{\mathcal{W}}_{\ell,m}(v,e)$ and $8\ell m \leq n^{1/3}$,*

$$|w(\gamma)| \leq c^{m+\chi} \left(\frac{1}{n}\right)^e \left(\frac{12\ell m}{\sqrt{n}}\right)^{(e_1 - 4\chi - 8m)_+}$$

with $\chi = e - v + 2$, and e_1 is the number of edges of $\boldsymbol{E}_\gamma$ with multiplicity $(1,1)$. Moreover,

$$e_1 \geq 2(e - \ell m).$$

Proof. Consider the graph $\widetilde{G}_\gamma$ defined in Lemma 38. Let $\widetilde{e}_1$ be the the number of edges of multiplicty 1 in $\widetilde{\gamma}$. Arguing as in the proof of Lemma 27, $\widetilde{e}_1 \geq 2(e - \ell m)$. The last statement is thus a consquence of the inequality $\tilde{e}_1 \leq e_1$. The first statement is a consequence Proposition 33 and the fact that the number of inconsistent edges of γ is at most in $4\chi + 8m$ (as already used in the proof of Lemma 38). □

All ingredients are in order to prove Proposition 35.

Proof of Proposition 35. For $n \geq 3$, we define

$$m = \left\lfloor \frac{\log n}{25 \log(\log n)} \right\rfloor .$$

We may then repeat the proof of Proposition 24 with the exponents slightly modified. ☐

5.4. *Proof of Theorem* 30. With Propositions 35 and 36, the net argument used in Section 4.5 to prove Theorem 17 can be applied exactly in the same way to prove Theorem 30.

6. **Proof of Theorems 3 and 5**

The proof of Theorems 3 and 5 has become standard in the last decade; it is based on the the linearization trick. Let us outline it here. Let $U = (U_1, \dots, U_d)$ be elements of a unital C^*-algebra A and $V = (V_1, \dots, V_d)$ be elements of a unital C^*-algebra B. Let P be a non-commutative polynomial in d free variables and their adjoints. Then the following are equivalent:

(i) For any P, $||P(U)|| = ||P(V)||$.

(ii) For any polynomial P with *matrix coefficients* with matrices of any size, $||P(U)|| = ||P(V)||$.

(iii) For any integer r, and $r \times r$ matrices $a_0, \dots, a_d$,

$$||a_0 \otimes 1 + a_1 \otimes U_1 + \cdots + a_d \otimes U_d|| = ||a_0 \otimes 1 + a_1 \otimes V_1 + \cdots + a_d \otimes V_d||.$$

(iv) For any integer r, and $r \times r$ matrices $a_0, \dots, a_d$ such that $a_0 \otimes 1 + a_1 \otimes U_1 + \cdots + a_d \otimes U_d$ and $a_0 \otimes 1 + a_1 \otimes V_1 + \cdots + a_d \otimes V_d$ are self-adjoint,

$$||a_0 \otimes 1 + a_1 \otimes U_1 + \cdots + a_d \otimes U_d|| = ||a_0 \otimes 1 + a_1 \otimes V_1 + \cdots + a_d \otimes V_d||.$$

Here, all norms are C^*-algebra norms. For an accessible proof, we refer to [32, p. 256], (Exercise 1 following Proposition 4). Note that in the case of unitary matrices (the case of interest to us) an essentially sufficient version was already proved by Pisier in [36].

Back to our restricted permutation matrices $(S_i)_{|\mathbf{1}^\perp}$ and $(S_i \otimes S_i)_{|V^\perp}$, the important point is that instead of treating general non-commuting polynomials in $(S_i)_{|\mathbf{1}^\perp}$ (resp. $(S_i \otimes S_i)_{|V^\perp}$) as in (i) and (ii), we treat degree one self-adjoint polynomials with matrix values as in (iv). In Theorems 2 and 4, we prove the asymptotic convergence of operator norms of operators of the form (1) for any integer r, and $r \times r$ matrices $a_0, \dots, a_d$. Therefore, by the above criterion (iv), it implies the result for any polynomial P in $(S_i)_{|\mathbf{1}^\perp}$ (resp. $(S_i \otimes S_i)_{|V^\perp}$) with matrix coefficients, as in (ii).

7. Proofs of auxiliary results

7.1. *Proof of Proposition 7*. We note that if the matrices a_i have non-negative entries, then an analog of the Alon-Boppana lower bound (6) holds in this context; see [12]. The proof of Proposition 7 is based on the notion of spectral measure. If A is of the form (1) and the symmetry condition (3) holds, then A is a bounded self-adjoint operator. If $\phi \in \mathbb{C}^r \otimes \ell^2(X)$, we denote by μ_A^ϕ the spectral measure of A; that is,

$$\mu_A^\phi(\cdot) = \langle \phi, E(\cdot)\phi \rangle,$$

where E is the spectral resolution of the identity of A. We have for any integer $k \geq 0$,

$$\int \lambda^k d\mu_A^\phi(\lambda) = \langle \phi, A^k \phi \rangle. \tag{73}$$

If $x \in X$, we also define the spectral measure

$$\mu_A^x(\cdot) = \frac{1}{r}\mathrm{tr}\{E(\cdot)_{xx}\} = \frac{1}{r}\sum_{j=1}^r \mu_A^{f_j \otimes \delta_x}(\cdot),$$

where we used the notation (11) for $E(\cdot)_{xx}$ and $(f_1, \dots, f_r)$ is an orthonormal basis of $\mathbb{C}^r$. Moreover, if $X = [n]$ is finite and $(\psi_1, \dots, \psi_{nr})$ is an orthonormal basis of eigenvectors of A with eigenvalues $\lambda_1, \dots, \lambda_{nr}$, we have

$$\mu_A^x = \frac{1}{r}\sum_{k=1}^{nr} \|(\psi_k)(x)\|_2^2 \delta_{\lambda_k},$$

where $\psi(x) \in \mathbb{C}^r$ is the projection of ψ on $\mathbb{C}^r \otimes \{\delta_x\}$. Recall the orthogonal decomposition of $H_0 \oplus H_1$ of $\mathbb{C}^r \otimes \ell^2(X)$. Let us assume without loss of generality that $\psi_1, \dots, \psi_r$ is an orthonormal basis of H_1. Then, for $1 \leq k \leq r$, $\psi_k = (f_k \otimes \mathbf{1})/\sqrt{n}$, where $(f_1, \dots, f_r)$ is an orthonormal basis of $\mathbb{C}^r$. We set

$$\mu_{A_{|H_0}}^x = \frac{1}{r}\sum_{k=r+1}^{nr} \|(\psi_k)(x)\|_2^2 \delta_{\lambda_k}.$$

It follows that

$$|\mu_A^x - \mu_{A_{|H_0}}^x|(\mathbb{R}) = \frac{1}{r}\sum_{k=1}^{r} \|(\psi_k)(x)\|_2^2 = \frac{1}{n},$$

and by construction,

$$\mathrm{supp}(\mu_{A_{|H_0}}^x) \subset \sigma(A_{|H_0}).$$

We readily deduce the following lemma.

LEMMA 40. *Assume that $X = [n]$ and that the symmetry condition (3) holds. If $f : \mathbb{R} \to \mathbb{R}$ is a function uniformly bounded by 1 and $\int f d\mu_A^x > 1/n$, then $\sigma(A_{|H_0}) \cap \mathrm{supp}(f) \neq 0$.*

We may now prove Proposition 7.

Proof of Proposition 7. Since $\sigma(A_\star)$ is compact, it is sufficient to prove that for any $\lambda \in \sigma(A_\star)$, there exists an integer $h = h(\varepsilon, \lambda)$ such that if $(G^\sigma, x)_h$ contains no cycle for some $x \in [n]$, then $\sigma(A_{|H_0}) \cap [\lambda - \varepsilon, \lambda + \varepsilon] \neq 0$.

Let $S = \{g_i : i \in [d]\}$ be the symmetric generating set of the free group $X_\star$, and let o be its unit. Let M be a bounded operator on $\mathbb{C}^r \otimes \ell^2(X_\star)$ in the $C^\star$-algebra generated by finite linear combinations of operators of the form $b \otimes \lambda(g)$. We introduce the standard tracial state τ defined in (18) (with r instead of k). Then

$$\int \lambda^k d\mu^o_{A_\star}(\lambda) = \tau(A^k_\star)$$

and

$$\operatorname{supp}(\mu^o_{A_\star}) = \sigma(A_\star).$$

In particular, if $\lambda \in \sigma(A_\star)$ and $f(x) = \max(0, 1 - |x - \lambda|/\varepsilon)$, we have

$$\eta = \int f \, d\mu^o_{A_\star} > 0.$$

Set $T = \sum_i \|a_i\|$ and $I = [-T, T]$. From the Stone-Weierstrass Theorem, there exists a polynomial p of degree m such that for any $x \in I$,

$$|f(x) - p(x)| < \eta/4.$$

Since the norms of A and $A_\star$ are bounded by T, we deduce that

$$\left| \int f \, d\mu^x_A - \int f \, d\mu^o_{A_\star} \right| < \left| \int p \, d\mu^x_A - \int p \, d\mu^o_{A_\star} \right| + \eta/2.$$

However, from (73), $\int \lambda^k \, d\mu^x_A(\lambda)$ is a function of $(G^\sigma, x)_h$. More precisely, we have

$$\int \lambda^k \, d\mu^x_A(\lambda) = \frac{1}{r} \sum_\gamma \operatorname{tr} \prod_{t=1}^k a_{i_t},$$

where the sum is over all closed paths $\gamma = (x_1, i_1, x_2, i_2, \ldots, x_{k+1})$ in G^σ of length k with $x_1 = x_{k+1} = x$. We deduce that if $(G^\sigma, x)_h$ contains no cycle, then for any integer $0 \leq k \leq 2h$,

$$\int \lambda^k \, d\mu^x_A(\lambda) = \int \lambda^k \, d\mu^o_{A_\star}(\lambda).$$

Hence, if $2h \geq m$, we obtain

$$\int p \, d\mu^x_A = \int p \, d\mu^o_{A_\star},$$

and consequently

$$\int f \, d\mu^x_A > \int f \, d\mu^o_{A_\star} - \left| \int p \, d\mu^x_A - \int p \, d\mu^o_{A_\star} \right| - \eta/2 > \eta/4.$$

If $n \geq 4/\eta$, then we may conclude using Lemma 40. (Note that the condition $n \geq \eta/4$ is included in the condition that $(G^\sigma, x)_{\lceil \eta/4 \rceil}$ contain no cycle.) □

7.2. *Proof of Theorem 16*.

Proof. Let us give a proof of (ii), which is the result that we have actually used. We only prove (i) when L itself is positive semi-definite. We may argue as in Gross [23, Th. 2]. From Lemma 15, L maps hermitian and skew-hermitian matrices to hermitian and skew-hermitian matrices. We deduce that if λ is a (non- negative) eigenvalue of L with eigenvector x, then both its hermitian and skew-hermitian parts are eigenvector. At least one of the two parts is non-trivial, and it follows that there exists an hermitian eigenvector x with eigenvalue ρ. We write $x = a - b$ with a, b positive semi-definite. We have $|x| = a + b$ and

$$\begin{aligned}\rho\langle |x|, |x|\rangle = \rho \mathrm{tr}(x^2) &= \langle Lx, x\rangle \\ &= \langle La, a\rangle + \langle Lb, b\rangle - 2\langle La, b\rangle \\ &\leq \langle La, a\rangle + \langle Lb, b\rangle + 2\langle La, b\rangle = \langle L|x|, |x|\rangle,\end{aligned}$$

where at the last line, we have used that if a, c are positive semi-definite, then $c^{1/2}ac^{1/2}$ is positive semi-definite and $\langle a, c\rangle = \mathrm{tr}(ac) = \mathrm{tr}(c^{1/2}ac^{1/2}) \geq 0$. Since ρ is the operator norm of L, we get

$$\rho\langle |x|, |x|\rangle \leq \langle L|x|, |x|\rangle \leq \rho\langle |x|, |x|\rangle.$$

Hence $\langle(\rho - L)|x|, |x|\rangle = 0$. Since $\rho - L$ is positive semi-definite, we thus have proved that $|x|$ is an eigenvector of L with eigenvalue ρ. It concludes the proof of (i) when L is positive semi-definite.

We may then prove (ii). From Lemma 15, $(L^n)^* L^n$ is positive semi-definite and of negative type. Let ρ_n^{2n} be the operator norm of $(L^n)^* L^n$. Gelfand's formula implies that ρ_n converges to ρ. Moreover, from what precedes, $(L^n)^* L^n$ has a positive semi-definite eigenvector y_n, with $\|y_n\|_2 = 1$ with eigenvalue ρ_n^{2n}. From the spectral theorem, we have

$$\rho_n^{2n}\|x\|_2^2 \geq \|L^n x\|_2^2 = \langle x, (L^n)^* L^n x\rangle \geq \rho_n^{2n}|\langle x, y_n\rangle|^2. \tag{74}$$

However, since for any positive semi-definite x, y, $\mathrm{tr}(xy) = \mathrm{tr}(y^{1/2}xy^{1/2}) \leq \mathrm{tr}(x)\|y\|$ (where $\|y\|$ is the operator norm), we deduce that

$$\begin{aligned}1 = \mathrm{tr}(y_n^2) \leq \mathrm{tr}(y_n) &= \mathrm{tr}(x^{-1/2}x^{1/2}y_n x^{1/2}x^{-1/2}) \\ &\leq \mathrm{tr}(x^{1/2}y_n x^{1/2})\|x^{-1}\| = \langle x, y_n\rangle\|x^{-1}\|.\end{aligned}$$

Hence $|\langle x, y_n\rangle|^2$ is lower bounded uniformly in n by some $\delta > 0$. Taking the power $1/(2n)$ in (74) concludes the proof of (ii). ∎

7.3. *Proof of Lemma* 3.4.

Proof. We adapt the proof of [9, Lemma 9]. Let $\varepsilon = (1-p-p/q)/(1-p)$, $\delta = -z\sqrt{q}$ and $f(x) = \mathbb{E}(-1)^N \prod_{n=0}^{2N-1}(1/\sqrt{x} - zn)$. We write

$$f(q) = \sum_{t=0}^{k} \binom{k}{t} p^t (1-p)^{k-t} (-1)^t \prod_{n=0}^{2t-1} \left(\frac{1}{\sqrt{q}} - zn \right)$$
$$= (1-p)^k \sum_{t=0}^{k} \binom{k}{t} (-1+\varepsilon)^t \prod_{n=0}^{2t-1} (1+\delta n).$$

We will use that by assumption, $p \le 1/4$ and $|\varepsilon| \le (4/3)(1/\sqrt{8}) \le 1/2$. We write

$$\prod_{n=0}^{2t-1} (1+\delta n) = 1 + \sum_{s=1}^{2t-1} \delta^s \sum_{(s)} \prod_{i=1}^{s} n_i = 1 + \sum_{s=1}^{2t-1} \delta^s P_s(2t),$$

where $\sum_{(s)}$ is the sum over all $(n_i)_{1\le i\le s}$ all distinct and $1 \le n_i \le 2t-1$. We observe that $t \mapsto P_s(t)$ is a polynomial of degree $2s$ in t, which vanishes at integers $0 \le t \le s$ and for integer $t \ge s+1$, $0 \le P_s(t) \le (\sum_{1\le n\le t-1} n)^s \le (t^2/2)^s$. Setting $P_0(t) = 1$, we have

$$|f(q)| \le \left| \sum_{s=0}^{2k-1} \delta^s \sum_{t=0}^{k} \binom{k}{t} (-1+\varepsilon)^t P_s(t) \right|. \tag{75}$$

We may then repeat verbatim the proof of [9, Lemma 9]. In (75), for large values of s, we have the rough bound

$$\sum_{s=\lfloor\frac{k-1}{2}\rfloor+1}^{2k-1} |\delta|^s \sum_{t=0}^{k} \binom{k}{t} |-1+\varepsilon|^t |P_s(2t)| \le \sum_{s=\lfloor\frac{k-1}{2}\rfloor+1}^{\infty} |\delta|^s 3^k k^{2s} 2^s$$
$$\le 2.3^k \left(2|\delta|k^2\right)^{\frac{k}{2}} = 2(3\sqrt{2}kz^{1/2}q^{1/4})^k,$$

where we have used that $|1-\varepsilon| \le 2$, $\sum_{t=0}^{k} \binom{k}{t} 2^t = 3^k$, $|P_s(2t)| \le (2k^2)^s$ and, at the third step, that $2|\delta|k^2 = 2zk^2\sqrt{q} \le 1/2$ and $\sum_{s\ge r} x^k \le 2x^r$ if $0 \le x \le 1/2$.

For $1 \le s \le \lfloor (k-1)/2 \rfloor$, there are some algebraic cancellations in (75). Consider the derivative of order m of $(1+x)^k = \sum_{t=0}^{k} \binom{k}{t} x^t$. It vanishes at $x = -1$ for any $0 \le m \le k-1$. We get that for any $0 \le m \le k-1$,

$$0 = \sum_{t=0}^{k} \binom{k}{t} (-1)^t (t)_m.$$

Since $Q_m(x) = (x)_m$ is a monic polynomial of degree m, $Q_0, \ldots, Q_{k-1}$ is a basis of $\mathbb{R}_{k-1}[x]$, the real polynomials of degree at most $k-1$. Hence, by linearity that for any $P \in \mathbb{R}_{k-1}[x]$,

$$0 = \sum_{t=0}^{k} \binom{k}{t} (-1)^t P(t). \tag{76}$$

Now, using again the binomial identity, we write

$$(1-\varepsilon)^t = \sum_{r=0}^{t} (-\varepsilon)^r \binom{t}{r} = T_{k,s}(t) + R_{k,s}(t),$$

where $T_{k,s}(t) = \sum_{r=0}^{k-1-2s} (-\varepsilon)^r \binom{t}{r}$ is a polynomial in t of degree $k-1-2s$. Using $0 \le \varepsilon \le 1/2$ and, if $t \ge 2$, using $0 \le r \le t$, $\binom{t}{r} \le 2^{t-1}$ (from Pascal's identity),

$$|R_{k,s}(t)| = \left| \sum_{r=k-2s}^{t} (-\varepsilon)^r \binom{t}{r} \right| \le 2^{t-1} \sum_{r=k-2s}^{t} \varepsilon^r \le 2^t \varepsilon^{k-2s}.$$

If $t \in \{0,1\}$, then this last inequality also holds. Since P_s is a polynomial of degree $2s$, we get from (76) that

$$I = \sum_{s=0}^{\lfloor \frac{k-1}{2} \rfloor} \delta^s \sum_{t=0}^{k} \binom{k}{t} (-1+\varepsilon)^t P_s(2t) = \sum_{s=0}^{\lfloor \frac{k-1}{2} \rfloor} \delta^s \sum_{t=0}^{k} \binom{k}{t} (-1)^t R_{k,s}(t) P_s(2t).$$

Taking absolute values and using again $|P_s(2t)| \le 2^s t^{2s}$, we find

$$\begin{aligned} |I| &\le \sum_{s=0}^{\lfloor \frac{k-1}{2} \rfloor} |\delta|^s \sum_{t=0}^{k} \binom{k}{t} 2^t \varepsilon^{k-2s} k^{2s} 2^s \\ &= (3\varepsilon)^k \sum_{s=0}^{\lfloor \frac{k-1}{2} \rfloor} \left(\frac{2|\delta| k^2}{\varepsilon^2} \right)^s \le 2(3\varepsilon)^k \left(\frac{2|\delta| k^2}{\varepsilon^2} \right)^{\frac{k}{2}}, \end{aligned}$$

where at the last step, we use that $|\delta| k^2/\varepsilon^2 \ge 1$ and $\sum_{s=0}^{r} x^s \le 2x^r$ if $x \ge 2$. We obtain $|I| \le 2(3\sqrt{2} k z^{1/2} q^{1/4})^k$. This concludes the proof of the lemma. □

References

[1] L. Addario-Berry and S. Griffiths, The spectrum of random lifts, 2010. arXiv 1012.4097.

[2] C. A. Akemann and P. A. Ostrand, Computing norms in group C^*-algebras, *Amer. J. Math.* **98** no. 4 (1976), 1015–1047. MR 0442698. Zbl 0342.22008. https://doi.org/10.2307/2374039.

[3] D. Aldous and R. Lyons, Processes on unimodular random networks, *Electron. J. Probab.* **12** (2007), no. 54, 1454–1508. MR 2354165. Zbl 1131.60003. https://doi.org/10.1214/EJP.v12-463.

[4] A. Amit and N. Linial, Random graph coverings. I. General theory and graph connectivity, *Combinatorica* **22** no. 1 (2002), 1–18. MR 1883559. Zbl 0996.05105. https://doi.org/10.1007/s004930200000.

[5] A. Amit and N. Linial, Random lifts of graphs: edge expansion, *Combin. Probab. Comput.* **15** no. 3 (2006), 317–332. MR 2216470. Zbl 1095.05034. https://doi.org/10.1017/S0963548305007273.

[6] N. Anantharaman, Quantum ergodicity on regular graphs, *Comm. Math. Phys.* **353** no. 2 (2017), 633–690. MR 3649482. Zbl 1368.58015. https://doi.org/10.1007/s00220-017-2879-9.

[7] G. W. Anderson, Convergence of the largest singular value of a polynomial in independent Wigner matrices, *Ann. Probab.* **41** no. 3B (2013), 2103–2181. MR 3098069. Zbl 1282.60007. https://doi.org/10.1214/11-AOP739.

[8] C. Bordenave, Spectrum of random graphs, in *Advanced Topics in Random Matrices*, *Panor. Synthèses* **53**, Soc. Math. France, Paris, 2017, pp. 91–150. MR 3792625. Zbl 1395.05149.

[9] C. Bordenave, A new proof of Friedman's second eigenvalue theorem and its extension to random lifts, *Ann. Sci. Éc. Norm. Supér.*, to appear.

[10] C. Bordenave, M. Lelarge, and L. Massoulié, Nonbacktracking spectrum of random graphs: community detection and nonregular Ramanujan graphs, *Ann. Probab.* **46** no. 1 (2018), 1–71. MR 3758726. Zbl 1386.05174. https://doi.org/10.1214/16-AOP1142.

[11] A. Buchholz, Norm of convolution by operator-valued functions on free groups, *Proc. Amer. Math. Soc.* **127** no. 6 (1999), 1671–1682. MR 1476122. Zbl 0916.43002. https://doi.org/10.1090/S0002-9939-99-04660-2.

[12] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli, Weighted expanders and the anisotropic Alon-Boppana theorem, *European J. Combin.* **25** no. 5 (2004), 735–744. MR 2061396. Zbl 1048.05057. https://doi.org/10.1016/j.ejc.2003.10.008.

[13] B. Collins and P. Y. Gaudreau Lamarre, $*$-freeness in finite tensor products, *Adv. in Appl. Math.* **83** (2017), 47–80. MR 3573218. Zbl 1369.46059. https://doi.org/10.1016/j.aam.2016.09.002.

[14] B. Collins and C. Male, The strong asymptotic freeness of Haar and deterministic matrices, *Ann. Sci. Éc. Norm. Supér.* (4) **47** no. 1 (2014), 147–163. MR 3205602. Zbl 1303.15043. https://doi.org/10.24033/asens.2211.

[15] K. Deimling, *Nonlinear Functional Analysis*, Springer-Verlag, Berlin, 1985. MR 0787404. Zbl 1257.47059. https://doi.org/10.1007/978-3-662-00547-7.

[16] A. Figà-Talamanca and T. Steger, Harmonic analysis for anisotropic random walks on homogeneous trees, *Mem. Amer. Math. Soc.* **110** no. 531 (1994), xii+68. MR 1219707. Zbl 0836.43019. https://doi.org/10.1090/memo/0531.

[17] J. Friedman, Relative expanders or weakly relatively Ramanujan graphs, *Duke Math. J.* **118** no. 1 (2003), 19–35. MR 1978881. Zbl 1035.05058. https://doi.org/10.1215/S0012-7094-03-11812-8.

[18] J. Friedman, A proof of Alon's second eigenvalue conjecture and related problems, *Mem. Amer. Math. Soc.* **195** no. 910 (2008), viii+100. MR 2437174. Zbl 1177.05070. https://doi.org/10.1090/memo/0910.

[19] J. Friedman, A. Joux, Y. Roichman, J. Stern, and J. P. Tillich, The action of a few random permutations on r-tuples and an application to cryptography, in *STACS* 96 (Grenoble, 1996), *Lecture Notes in Comput. Sci.* **1046**, Springer-Verlag, Berlin, 1996, pp. 375–386. MR 1462111. Zbl 1380.94089. https://doi.org/10.1007/3-540-60922-9_31.

[20] J. Friedman and D.-E. Kohler, The relativized second eigenvalue conjecture of alon, 2014. arXiv 1403.3462.

[21] Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrices, *Combinatorica* **1** no. 3 (1981), 233–241. MR 0637828. Zbl 0494.15010. https://doi.org/10.1007/BF02579329.

[22] Y. Greenberg, Spectra of graphs and their covering trees, 1995, Ph.D. thesis, Hebrew University of Jerusalem.

[23] L. Gross, Existence and uniqueness of physical ground states, *J. Functional Analysis* **10** (1972), 52–109. MR 0339722. Zbl 0237.47012. https://doi.org/10.1016/0022-1236(72)90057-2.

[24] U. Haagerup and S. Thorbjørnsen, A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group, *Ann. of Math.* (2) **162** no. 2 (2005), 711–775. MR 2183281. Zbl 1103.46032. https://doi.org/10.4007/annals.2005.162.711.

[25] M. B. Hastings, Random unitaries give quantum expanders, *Phys. Rev. A* (3) **76** no. 3 (2007), 032315, 11. MR 2486279. https://doi.org/10.1103/PhysRevA.76.032315.

[26] M. B. Hastings and A. W. Harrow, Classical and quantum tensor product expanders, *Quantum Inf. Comput.* **9** no. 3-4 (2009), 336–360. MR 2553116. Zbl 1171.81329. Available at http://www.rintonpress.com/xxqic9/qic-9-34/0336-0360.pdf.

[27] M. G. Kreĭn and M. A. Rutman, Linear operators leaving invariant a cone in a Banach space, *Amer. Math. Soc. Translation* **1950** no. 26 (1950), 128, Translated from Usp. Mat. Nauk (N.S.) **3**, no. 1(23), 3–95 (1948). MR 0038008. Zbl 0030.12902.

[28] F. Lehner, Computing norms of free operators with matrix coefficients, *Amer. J. Math.* **121** no. 3 (1999), 453–486. MR 1738412. Zbl 0929.22004. https://doi.org/10.1353/ajm.1999.0022.

[29] N. Linial and D. Puder, Word maps and spectra of random graph lifts, *Random Structures Algorithms* **37** no. 1 (2010), 100–135. MR 2674623. Zbl 1242.60011. https://doi.org/10.1002/rsa.20304.

[30] E. Lubetzky, B. Sudakov, and V. Vu, Spectra of lifted Ramanujan graphs, *Adv. Math.* **227** no. 4 (2011), 1612–1645. MR 2799807. Zbl 1222.05168. https://doi.org/10.1016/j.aim.2011.03.016.

[31] L. Massoulié, Community detection thresholds and the weak Ramanujan property, in *STOC*'14—*Proceedings of the* 2014 *ACM Symposium on Theory of Computing*, ACM, New York, 2014, pp. 694–703. MR 3238997. Zbl 1315.68210. https://doi.org/10.1145/2591796.2591857.

[32] J. A. Mingo and R. Speicher, *Free Probability and Random Matrices*, *Fields Institute Monographs* **35**, Springer, New York; Fields Institute for Research in

Mathematical Sciences, Toronto, ON, 2017. MR 3585560. Zbl 1387.60005. https://doi.org/10.1007/978-1-4939-6942-5.

[33] G. J. Murphy, C^**-Algebras and Operator Theory*, Academic Press, Inc., Boston, MA, 1990. MR 1074574. Zbl 0714.46041. https://doi.org/10.1016/C2009-0-22289-6.

[34] A. Nica, Asymptotically free families of random unitaries in symmetric groups, *Pacific J. Math.* **157** no. 2 (1993), 295–310. MR 1197059. Zbl 0739.46051. https://doi.org/10.2140/pjm.1993.157.295.

[35] R. I. Oliveira, The spectrum of random k-lifts of large graphs (with possibly large k), *J. Comb.* **1** no. 3-4 (2010), 285–306. MR 2799213. Zbl 1244.05147. https://doi.org/10.4310/JOC.2010.v1.n3.a2.

[36] G. Pisier, A simple proof of a theorem of Kirchberg and related results on C^*-norms, *J. Operator Theory* **35** no. 2 (1996), 317–335. MR 1401692. Zbl 0858.46045. Available at http://www.mathjournals.org/jot/1996-035-002/1996-035-002-005.pdf.

[37] G. Pisier, Quantum expanders and geometry of operator spaces, *J. Eur. Math. Soc.* (*JEMS*) **16** no. 6 (2014), 1183–1219. MR 3226740. Zbl 1307.46045. https://doi.org/10.4171/JEMS/458.

[38] D. Puder, Expansion of random graphs: new proofs, new results, *Invent. Math.* **201** no. 3 (2015), 845–908. MR 3385636. Zbl 1320.05115. https://doi.org/10.1007/s00222-014-0560-x.

[39] Y. Watanabe and K. Fukumizu, Graph zeta function in the Bethe free energy and Loopy Belief Propagation, in *Advances in Neural Information Processing Systems* 22, 2009, (Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors), pp. 2017–2025. Available at https://papers.nips.cc/paper/3779-graph-zeta-function-in-the-bethe-free-energy-and-loopy-belief-propagation.

(Received: January 20, 2018)
(Revised: April 30, 2019)

Institut Mathématiques de Marseille, Marseille, France
E-mail: charles.bordenave@univ-amu.fr

Department of Mathematics, Kyoto University, Kyoto, Japan
E-mail: collins@math.kyoto-u.ac.jp